

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE
OF MATHEMATICAL STATISTICS

Contents

	PAGE
Parabolic Test for Linkage. N. L. JOHNSON.....	227
Reduction of a Certain Class of Composite Statistical Hypotheses. GEORGE W. BROWN.....	254
The Selection of Variates for Use in Prediction with some Com- ments on the Problem of Nuisance Parameters. HAROLD HOTELLING.....	271
The Fitting of Straight Lines if Both Variables are Subject to Error. ABRAHAM WALD.....	284
A Method for Minimizing the Sum of Absolute Values of Devia- tions. ROBERT R. SINGLETON.....	301
A Study of a Universe of n Finite Populations with Application to Moment-Function Adjustments for Grouped Data. JOSEPH A. PIERCE.....	311
The Analysis of Variance when Experimental Errors follow the Poisson or Binomial Laws. W. G. COCHRAN.....	335
Notes:	
Orthogonal Polynomials Applied to Least Square Fitting of Weighted Observations. BRADFORD F. KIMBALL.....	348
Combinatorial Formulas for the r^{th} Standard Moment of the Sample Sum, of the Sample Mean, and of the Normal Curve. P. S. DWYER.	353
On a Method of Sampling. E. G. OLDS.....	355
Rank Correlation when there are Equal Variates. MAX A. WOODBURY.	358
Note on Theoretical and Observed Distributions of Repetitive Occur- rences. P. S. OLMSTEAD.....	363

THE ANNALS OF MATHEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

A. T. CRAIG

J. NEYMAN

WITH THE COÖPERATION OF

H. C. CARVER

R. A. FISHER

R. DE MISES

H. CRAMÉR

T. C. FRY

E. S. PEARSON,

W. E. DEMING

H. HOTELLING

H. L. RIETZ

G. DARMOIS

W. A. SHEWHART

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, P. R. Rider, Washington University, St. Louis, Mo.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is \$4.00 per year. Single copies \$1.25. Back numbers are available at the following rates:

Vols. I-IV \$5.00 each. Single numbers \$1.50.

Vols. V to date \$4.00 each. Single numbers \$1.25.

PARABOLIC TEST FOR LINKAGE

BY N. L. JOHNSON

1. Introduction. In this paper a problem in testing statistical hypotheses which has applications in genetics will be treated from the standpoint of the Neyman-Pearson approach. This approach has been developed in a series of papers, [4], [5], [6], [7], [8], [9], [10], to which the reader is referred for definitions of the concepts of a simple statistical hypothesis, critical regions, power function of a test with respect to alternative hypotheses, and that of a test unbiased in the limit employed in the present paper.

2. Statement of Problem. We shall consider M independent experiments, which will each yield results falling into one of the four categories described by the possible combinations of the 4 events a , not- a (or \bar{a}), b , and not- b (or \bar{b}) as set up in the following table.

	a	not- a	
b	p_1	p_2	P_1
not- b	p_3	p_4	$1 - P_1$
	P_2	$1 - P_2$	1

We shall assume that the marginal probabilities are known and have values $P_1, 1 - P_1, P_2, 1 - P_2$ as shown in the table. Thus P_1 = probability of event b happening whether event a occurs or not. It is obvious that if, further, the probability of a result falling in any one category or cell is fixed, then the other three cell probabilities will also be fixed. For if p_1, p_2, p_3, p_4 be the four cell probabilities as shown in the table above, we must have

$$(1) \quad p_1 + p_2 = P_1; \quad p_1 + p_3 = P_2; \quad p_2 + p_4 = 1 - P_2.$$

Hence the values of the cell probabilities will be determined by a single parameter θ , say, as follows

$$(2) \quad \begin{aligned} p_1 &= P_1 P_2 e^\theta & p_2 &= P_1 (1 - P_2 e^\theta) \\ p_3 &= P_2 (1 - P_1 e^\theta) & p_4 &= 1 - P_1 - P_2 + P_1 P_2 e^\theta. \end{aligned}$$

The range of values which θ may take for the set of admissible hypotheses is found from the conditions

$$(3) \quad 0 \leq p_i \leq 1 \quad (i = 1, 2, 3, 4)$$

to be

$$(4) \quad -\infty < \theta \leq \min(-\log P_1, -\log P_2) \quad \text{if } P_1 + P_2 \leq 1$$

but

$$(5) \quad \log(P_1^{-1} + P_2^{-1} - P_1^{-1}P_2^{-1}) \leq \theta \leq \min(-\log P_1, -\log P_2) \quad \text{if } P_1 + P_2 \geq 1.$$

The hypothesis tested, H_0 , is that $\theta = 0$, i.e. that the events a and b are independent. It will be noticed that H_0 is a simple hypothesis, since it specifies the probability law of the observed variables completely. In fact, if m_i be the number of results out of our M experiments which are in the i th category, then m_1, m_2, m_3, m_4 are our observed variables, and we have

$$(6) \quad P\{m_1 = m'_1, m_2 = m'_2, m_3 = m'_3, m_4 = m'_4 \mid H_0\} = \frac{M! p_{01}^{m'_1} p_{02}^{m'_2} p_{03}^{m'_3} p_{04}^{m'_4}}{m'_1! m'_2! m'_3! m'_4!}$$

where p_{0i} is the value of p_i when $\theta = 0$.

This is the conceptual model used in testing for linkage in two pairs of genes; H_0 corresponds to the hypothesis "there is no linkage." Fuller explanations are given by Fisher [3]. It should be noted, however, that Fisher uses a parameter θ corresponding to $\frac{1}{4}e^{\theta}$ in this paper.

3. Basis of Selection of Test. The question now arises; what test shall we choose for the hypothesis H_0 ? That is, what should the critical region w be to give us results as satisfactory as possible? The main aim must be to avoid errors, both of first and second kind, as far as possible. The first kind of error is subject to control, since the probability of the sample point E falling in w when H_0 is true (which we shall denote by $P\{E \in w \mid H_0\}$) can be determined approximately, H_0 being simple. The critical region w is therefore chosen, if possible, to give a definite level of significance to the test associated with it. However, there will usually be many regions which will do this, and in order to decide which of them give more satisfactory results we consider $(1 - P\{E \in w \mid H\})$; i.e. the probability of the second kind of error with respect to an alternative hypothesis H , the first kind of error being fixed.

In the present case H will be determined by θ and so we may put $P\{E \in w \mid H\} = \beta(w \mid \theta)$, where $\beta(w \mid \theta)$, considered as a function of θ , will be the power function of the test associated with the critical region w . We want w to be such that $\beta(w \mid 0) = \alpha$. α being the fixed level of significance while $\beta(w \mid \theta)$ is as large as possible.

It is also desirable that we should accept the hypothesis H_0 more often when it is true than when any one of the alternative hypotheses (H) is true. Ex-

pressed symbolically, this means that

$$(7) \quad \beta(w | 0) \leq \beta(w | \theta) \quad \text{for all } \theta \neq 0.$$

Any test satisfying the last condition is said to be *unbiased*.

If β and $\frac{\partial \beta}{\partial \theta}$ are each continuous and differentiable functions of θ , and we consider only those alternative hypotheses specified by suitably small values of θ , sufficient conditions for the test to be unbiased will be

$$(8) \quad \left. \frac{\partial \beta}{\partial \theta} \right|_{\theta=0} = 0,$$

$$(9) \quad \left. \frac{\partial^2 \beta}{\partial \theta^2} \right|_{\theta=0} > 0.$$

According to the terminology recently adopted by Daly [1], the tests of which it is known only that they satisfy (8) and (9), are called *locally unbiased*.

If a region w could be found such that, v being any other region for which

$$(10) \quad \beta(w | 0) = \beta(v | 0), \quad \text{then } \beta(w | \theta) \geq \beta(v | \theta)$$

for all $\theta \neq 0$, this would give a test which would be the best with respect to any alternative hypothesis. However, it has been shown by Neyman [4] that under certain conditions, which many probability laws satisfy, such a test will not exist. An attempt is therefore made to control the power of the test with respect to hypotheses specifying values of θ near to 0; hoping that the powers of the tests so obtained with respect to the other hypotheses will behave in a satisfactory manner. Thus Neyman and Pearson [9] define an "unbiased test of Type A" as a test corresponding to a critical region w such that if v be any other region in the sample space W for which

$$(11) \quad \beta(w | 0) = \beta(v | 0) = \alpha$$

and

$$(12) \quad \left. \frac{\partial \beta(w | \theta)}{\partial \theta} \right|_{\theta=0} = \left. \frac{\partial \beta(v | \theta)}{\partial \theta} \right|_{\theta=0} = 0$$

then

$$(13) \quad \left. \frac{\partial^2 \beta(w | \theta)}{\partial \theta^2} \right|_{\theta=0} \geq \left. \frac{\partial^2 \beta(v | \theta)}{\partial \theta^2} \right|_{\theta=0}.$$

In the problem which I am treating the conditions

$$(14) \quad \beta(w | 0) = \alpha; \quad \left. \frac{\partial \beta(w | \theta)}{\partial \theta} \right|_{\theta=0} = 0$$

implied by (11) and (12) above cannot, in general, be satisfied, since the distribution is discontinuous, i.e. $P\{E \in w | H_0\}$ is a discontinuous function of w and, in

fact, for a given sample size, has only a finite number of possible values, none of which need be equal to α .

However, it may be possible to find a test of H_0 of a type called "unbiased in the limit (as M increases)," based on the limiting form of the multinomial distribution which is a continuous function of w . The definition [6] of a test "unbiased in the limit" will be taken as follows:

Suppose we have a sequence (w_M) of critical regions, w_M corresponding to a sample of size M , such that

(i) for any M , if v_M be any region for which

$$(15) \quad \beta(w_M | 0) = \beta(v_M | 0)$$

and

$$(16) \quad \left. \frac{\partial \beta(w_M | \theta)}{\partial \theta} \right|_{\theta=0} = \left. \frac{\partial \beta(v_M | \theta)}{\partial \theta} \right|_{\theta=0}$$

then

$$(17) \quad \left. \frac{\partial^2 \beta(w_M | \theta)}{\partial \theta^2} \right|_{\theta=0} \geq \left. \frac{\partial^2 \beta(v_M | \theta)}{\partial \theta^2} \right|_{\theta=0}$$

(ii)

$$(18) \quad \lim_{M \rightarrow \infty} \beta(w_M | 0) = \alpha,$$

(iii) if

$$(19) \quad \vartheta = \sqrt{M} (\theta - 0) = \sqrt{M} \theta$$

$$(20) \quad \lim_{M \rightarrow \infty} \left. \frac{\partial \beta(w_M | \vartheta)}{\partial \vartheta} \right|_{\vartheta=0} = 0$$

then the test associated with this sequence of critical regions is unbiased in the limit. I shall call such a test a test of type A_∞ .

The reason for using ϑ as the variable in condition (19) above is that, unless our sequence of critical regions has been very badly or unluckily chosen, we shall have

$$(21) \quad \lim_{M \rightarrow \infty} \beta(w_M | \theta) = 1 \quad (\theta \neq 0)$$

while, by (18), $\lim_{M \rightarrow \infty} \beta(w_M | 0) = \alpha$ and so, in general, $\lim_{M \rightarrow \infty} \frac{\partial \beta(w_M | \theta)}{\partial \theta}$ will not exist at $\theta = 0$. Hence we introduce ϑ , termed the *normalized error*, and, keeping ϑ constant (and hence making θ tend to zero) we form $\lim_{M \rightarrow \infty} \frac{\partial \beta(w_M | \vartheta)}{\partial \vartheta}$.

In the next section will be obtained a test of H_0 which is of type A_∞ .

4. Derivation of Test. The composition of a sample of M experiments is uniquely determined by the numbers of results m_1, m_2, m_3 falling in the 1st,

2nd and 3rd categories respectively. Thus any sample may be represented by a point $E(m)$ in a three-dimensional sample space $W(m)$ with coordinate axes of m_1, m_2 , and m_3 . It will occasionally be convenient to represent the sample by a point in a three-dimensional space with other axes. The following sample spaces will be used.

$W(m)$ —space with coordinate axes of m_1, m_2, m_3

$W(d)$ — “ “ “ “ “ d_1, d_2, d_3

$W(x)$ — “ “ “ “ “ x_1, x_2, x_3

$W(n)$ — “ “ “ “ “ n_1, n_2, n_3

where

$$(22) \quad d_i = m_i - Mp_{0i} \quad (i = 1, 2, 3, 4)$$

$$(23) \quad x_i = (m_i - Mp_{0i})/(Mp_{0i})^{\frac{1}{2}} \quad (i = 1, 2, 3, 4)$$

$$(24) \quad n_i = m_i/M \quad (i = 1, 2, 3, 4).$$

I shall use w_M indifferently to denote “the critical region corresponding to sample size M ” in any of the four sample spaces above; E indifferently to denote corresponding positions of the sample point in any of the four sample spaces: except in cases where confusion might arise, where I shall use $w_M(m)$, $w_M(d)$, $w_M(x)$, $w_M(n)$ and $E(m)$, $E(d)$, $E(x)$, $E(n)$. When necessary the size of sample with which a point E is associated will be denoted by a subscript; e.g. E_M .

In finding a test of type A_∞ we shall need to consider the quantities $\beta(w_M | 0)$, $\left. \frac{\partial \beta(w_M | \vartheta)}{\partial \vartheta} \right|_{\vartheta=0}$, and $\left. \frac{\partial^2 \beta(w_M | \theta)}{\partial \theta^2} \right|_{\theta=0}$, where $\vartheta = \theta \sqrt{M}$.

The probability law of the observed values m_1, m_2, m_3 is discontinuous with respect to the points of the sample space W_M . For if E^0 be a point which corresponds to integral values m_1^0, m_2^0, m_3^0 of m_1, m_2, m_3 ; subject to the restrictions

$$(25) \quad 0 \leq m_i^0 \quad (i = 1, 2, 3)$$

$$(26) \quad 0 \leq \sum_{i=1}^3 m_i^0 \leq M$$

then

$$(27) \quad P\{E_M \equiv E^0 | \theta = 0\} = \frac{M! p_{01}^{m_1^0} p_{02}^{m_2^0} p_{03}^{m_3^0} p_{04}^{m_4^0}}{m_1^0! m_2^0! m_3^0! m_4^0!}$$

where

$$(28) \quad \sum_{i=1}^4 m_i^0 = M$$

and

$$(29) \quad \begin{aligned} p_{01} &= P_1 P_2 & p_{02} &= P_1(1 - P_2) \\ p_{03} &= P_2(1 - P_1) & p_{04} &= (1 - P_1)(1 - P_2) \end{aligned}$$

while if E^0 be not such a point

$$(30) \quad P\{E_M \equiv E^0 \mid \theta\} = 0$$

whatever the value of θ may be. Now

$$(31) \quad \beta(w_M \mid \theta) = \sum_{w_M} \frac{M! p_1^{m_1} p_2^{m_2} p_3^{m_3} p_4^{m_4}}{m_1! m_2! m_3! m_4!}$$

where p_1, p_2, p_3, p_4 are as defined in (2) above, and \sum_{w_M} denotes a finite summation over all points E' in w_M for which $P\{E_M \equiv E' \mid \theta\} \neq 0$. Differentiating each side of (31) with respect to θ , we get

$$(32) \quad \left. \frac{\partial \beta(w_M \mid \theta)}{\partial \theta} \right]_{\theta=0} = \sum_{w_M} \frac{M! p_{01}^{m_1} p_{02}^{m_2} p_{03}^{m_3} p_{04}^{m_4}}{m_1! m_2! m_3! m_4!} \times \left[\frac{m_1(1 - P_1 - P_2) - m_2 P_2 - m_3 P_1 + m_4 P_1 P_2}{(1 - P_1)(1 - P_2)} \right]$$

and

$$(33) \quad \left. \frac{\partial^2 \beta(w_M \mid \theta)}{\partial \theta^2} \right]_{\theta=0} = \sum_{w_M} \frac{M! p_{01}^{m_1} p_{02}^{m_2} p_{03}^{m_3} p_{04}^{m_4}}{m_1! m_2! m_3! m_4!} \cdot \frac{1}{(1 - P_1)^2(1 - P_2)^2} [\{m_1(1 - P_1 - P_2) - m_2 P_2 - m_3 P_1 + M P_1 P_2\}^2 - \{m_1 P_1 P_2(1 - P_1 - P_2) + m_2 P_2(1 - P_1 - P_1 P_2) + m_3 P_1(1 - P_2 - P_1 P_2) - M P_1 P_2(1 - P_1)(1 - P_2)\}].$$

THEOREM 1. The sequence of critical regions (w_M) defined by

$$(34) \quad v + B u^2 \geq A \text{ in } w_M; \quad v + B u^2 < A \text{ elsewhere,}$$

where

$$(35) \quad u = \frac{x_1(P_1 P_2)^{\frac{1}{2}}(1 - P_1 - P_2) - x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}} P_2 - x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}} P_1}{\{P_1 P_2(1 - P_1)(1 - P_2)\}^{\frac{1}{2}}}$$

$$(36) \quad v = \frac{P_1(1 - P_1)(2P_2 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}}\} + P_2(1 - P_2)(2P_1 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}}\}}{[P_1 P_2(1 - P_1)(1 - P_2)\{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2\}]^{\frac{1}{2}}}$$

$$(37) \quad B = \left[\frac{M P_1 P_2(1 - P_1)(1 - P_2)}{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2} \right]^{\frac{1}{2}}$$

$$(43) \quad |\beta(w|0) - I(w)| < \epsilon$$

where

$$(44) \quad I(w) = \frac{1}{(2\pi)^{\frac{1}{2}} \sqrt{p_{04}}} \iiint_{w(x)} e^{-\frac{1}{2}x_0^2} dx_1 dx_2 dx_3$$

and

$$(45) \quad x_0^2 = \sum_{i=1}^3 x_i^2 (1 + p_{0i} p_{04}^{-1}) + 2 \sum_{i < j \leq 3} x_i x_j (p_{0i} p_{0j})^{\frac{1}{2}} p_{04}^{-1}.$$

We will now apply a transformation to the coordinates m_1, m_2, m_3 which will

(a) transform inequality (42) into a simpler form,

(b) transform $I(w)$ into a form to which the tables of the Normal Probability Integral may easily be applied for purposes of calculation.

This transformation is

$$(46) \quad u = \frac{x_1(P_1 P_2)^{\frac{1}{2}}(1 - P_1 - P_2) - x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}} P_2 - x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}} P_1}{\{P_1 P_2(1 - P_1)(1 - P_2)\}^{\frac{1}{2}}}$$

$$(47) \quad v = \frac{P_1(1 - P_1)(2P_2 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}}\} + P_2(1 - P_2)(2P_1 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}}\}}{[P_1 P_2(1 - P_1)(1 - P_2)\{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2\}]^{\frac{1}{2}}}$$

$$(48) \quad t = \frac{(2P_1 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_3 P_2^{\frac{1}{2}}(1 - P_1)^{\frac{1}{2}}\} - (2P_2 - 1)\{x_1(P_1 P_2)^{\frac{1}{2}} + x_2 P_1^{\frac{1}{2}}(1 - P_2)^{\frac{1}{2}}\}}{\{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2\}^{\frac{1}{2}}}.$$

This is a proper transformation, since under the conditions of the theorem $0 < P_i < 1$ and P_1 and P_2 are not both $\frac{1}{2}$; and the Jacobian

$$(49) \quad J = \frac{\partial(u, v, t)}{\partial(x_1, x_2, x_3)} = p_{04}^{-1}$$

is non-zero and of constant sign.

Also

$$(50) \quad x_0^2 = u^2 + v^2 + t^2.$$

Hence

$$(51) \quad I(w) = \frac{1}{(2\pi)^{\frac{1}{2}}} \iiint_{w(u,v,t)} e^{-\frac{1}{2}(u^2+v^2+t^2)} du dv dt.$$

The inequality (42) is transformed into an inequality of form $B(u - a_6)^2 + v \geq A$ where B has the value stated above; a_6 and A being at present arbitrary constants.

Therefore we may put $a_5 = 0$ and define A by the equation

$$(52) \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha$$

and conclude that the sequence of critical regions (w_M) defined by the inequalities

$$(53) \quad \begin{aligned} Bu^2 + v &\geq A \quad \text{in } w_M \\ Bu^2 + v &< A \quad \text{elsewhere} \end{aligned}$$

will satisfy conditions (i) for a test of type A_∞ .

From (51) and (52)

$$(54) \quad \begin{aligned} I(w_M) &= \frac{1}{(2\pi)^3} \iiint_{w_M} e^{-\frac{1}{2}(u^2+v^2+t^2)} du dv dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha. \end{aligned}$$

By THEOREM 1 of the appendix, as mentioned above, we have

$$(55) \quad |\beta(w_M | 0) - I(w_M)| < \epsilon \quad \text{for all } M > M_\epsilon$$

i.e.

$$(56) \quad |\beta(w_M | 0) - \alpha| < \epsilon \quad \text{for all } M > M_\epsilon$$

and so

$$(57) \quad \beta(w_M | 0) \rightarrow \alpha \quad \text{as } M \rightarrow \infty.$$

Thus the sequence of critical regions (w_M) satisfies the condition (ii) of the definition of a test of type A_∞ .

If w be any region defined by inequalities on u and v only (as are the regions w_M) then, as a special case of THEOREM 1 of the Appendix, we have that for any $\epsilon > 0$ there exists a number M_ϵ such that for all $M > M_\epsilon$

$$(58) \quad \left| P_M(w) - \frac{1}{2\pi} \iint_{w(u,v)} e^{-\frac{1}{2}(u^2+v^2)} du dv \right| < \epsilon$$

where $P_M(w) = P\{E_M \in w | 0\}$.

By (31) and (32), noting that $\frac{\partial \beta(w | \theta)}{\partial \theta} = \sqrt{M} \cdot \frac{\partial \beta(w | \vartheta)}{\partial \vartheta}$, we have

$$(59) \quad \begin{aligned} \left. \frac{\partial \beta(w | \vartheta)}{\partial \vartheta} \right]_{\vartheta=0} &= \sum_w f_1(u, v) \cdot u \cdot (P_1 P_2)^{\frac{1}{2}} (1 - P_1)^{-\frac{1}{2}} (1 - P_2)^{-\frac{1}{2}} \\ &= \sum_w f_1(u, v) \cdot uk \end{aligned}$$

where $k = (P_1 P_2)^{\frac{1}{2}} (1 - P_1)^{-\frac{1}{2}} (1 - P_2)^{-\frac{1}{2}} > 0$.

By THEOREM 1 of the Appendix, as last stated above, we have

$$(60) \quad f_1(u, v) = \frac{1}{2\pi} \Delta u \Delta v \cdot e^{-\frac{1}{2}(u^2+v^2)} (1 + R_M)$$

where for convenience we have written Δu , Δv for $\Delta_{(M)} u$, $\Delta_{(M)} v$ the units of u and v when sample size is M , and R_M for $R_M(u, v)$ which has the property that

$$(61) \quad \sum_w R_M(u, v) \Delta_{(M)} u \cdot \Delta_{(M)} v \cdot e^{-\frac{1}{2}(u^2+v^2)} \rightarrow 0$$

uniformly with respect to w as $M \rightarrow \infty$.

Now let w^+ denote that part of w where $R_M \geq 0$ and w^- that part of w where $R_M < 0$. Then

$$(62) \quad \sum_{w^+} k u f_1(u, v) = \sum_{w^+} k \cdot \frac{\Delta u \Delta v}{2\pi} \cdot u e^{-\frac{1}{2}(u^2+v^2)} + \sum_{w^+} k \frac{\Delta u \Delta v}{2\pi} u R_M e^{-\frac{1}{2}(u^2+v^2)}.$$

Let

$$(63) \quad \begin{aligned} S_M^+ &= \sum_{w^+} k \frac{\Delta u \Delta v}{2\pi} u R_M e^{-\frac{1}{2}(u^2+v^2)} \\ &= k \sum_{w^+} \left\{ \left(R_M \frac{\Delta u \Delta v}{2\pi} \right)^{\frac{1}{2}} u e^{-\frac{1}{2}(u^2+v^2)} \right\} \left\{ \left(R_M \frac{\Delta u \Delta v}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{1}{2}(u^2+v^2)} \right\}. \end{aligned}$$

By Schwarz's inequality

$$(64) \quad \left| \frac{S_M^+}{k} \right| \leq \left| \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 R_M e^{-\frac{1}{2}(u^2+v^2)} \right|^{\frac{1}{2}} \left| \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} R_M e^{-\frac{1}{2}(u^2+v^2)} \right|^{\frac{1}{2}}.$$

But

$$(65) \quad \sum_{w^+} u^2 f_1(u, v) = \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 e^{-\frac{1}{2}(u^2+v^2)} + \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 R_M e^{-\frac{1}{2}(u^2+v^2)}.$$

Now $u^2 f_1(u, v) \geq 0$ and $\sum_w u^2 f_1(u, v)$ is finite (since u^2 is a homogeneous function of second degree in the x_i 's and so has a finite expectation) and is bounded as $M \rightarrow \infty$. Hence $\sum_{w^+} u^2 f_1(u, v)$ is finite and bounded as $M \rightarrow \infty$. Further, as $M \rightarrow \infty$

$$(66) \quad \sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 e^{-\frac{1}{2}(u^2+v^2)} \rightarrow \frac{1}{2\pi} \iint_{w^+} u^2 e^{-\frac{1}{2}(u^2+v^2)} du dv.$$

Hence $\sum_{w^+} \frac{\Delta u \Delta v}{2\pi} u^2 R_M e^{-\frac{1}{2}(u^2+v^2)}$ is bounded as $M \rightarrow \infty$. From this result, together with (61) and (64) it follows that $S_M^+ \rightarrow 0$ as $M \rightarrow \infty$ uniformly with respect to w . Putting

$$(67) \quad S_M^- = \sum_{w^-} k \frac{\Delta u \Delta v}{2\pi} u R_M e^{-\frac{1}{2}(u^2+v^2)}$$

it will follow in a similar manner that $S_M^- \rightarrow 0$ as $M \rightarrow \infty$ uniformly with respect to w . Hence

$$(68) \quad \begin{aligned} \left. \frac{\partial \beta(w | \vartheta)}{\partial \vartheta} \right]_{\vartheta=0} &= \sum_w k u f_1(u, v) \\ &= \sum_w k \frac{\Delta u \Delta v}{2\pi} u e^{-\frac{1}{2}(u^2+v^2)} + S_M \end{aligned}$$

where $S_M = S_M^+ + S_M^-$ and so $S_M \rightarrow 0$ as $M \rightarrow \infty$ uniformly with respect to w .

Hence whatever be $\epsilon > 0$, there is a number M' such that for all $M > M'$

$$(69) \quad \left| \frac{\partial \beta(w | \vartheta)}{\partial \vartheta} \right]_{\vartheta=0} - \frac{k}{2\pi} \iint_w u e^{-\frac{1}{2}(u^2+v^2)} du dv \right| < \epsilon$$

whatever be the region w . In particular we may take $w \equiv w_M$, and then we have

$$(70) \quad \frac{k}{2\pi} \iint_{w_M} u e^{-\frac{1}{2}(u^2+v^2)} du dv = \frac{k}{2\pi} \int_{-\infty}^{+\infty} \left\{ u e^{-\frac{1}{2}u^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = 0$$

and so

$$(71) \quad \left| \frac{\partial \beta(w_M | \vartheta)}{\partial \vartheta} \right]_{\vartheta=0} \right| < \epsilon \quad \text{for all } M > M'$$

i.e.,

$$(72) \quad \lim_{M \rightarrow \infty} \frac{\partial \beta(w_M | \vartheta)}{\partial \vartheta} \Big|_{\vartheta=0} = 0.$$

Hence the sequence of critical regions (w_M) satisfies condition (iii) for a test of type A_{∞} . This completes the proof of THEOREM 1.

In the above theorem we have found a test which is unbiased in the limit for all cases except that for which $P_1 = P_2 = \frac{1}{2}$. The following theorem derives the test appropriate to this special case, and it is found that in this instance the test takes a very simple form.

THEOREM 2. If $P_1 = P_2 = \frac{1}{2}$, the sequence of critical regions (w_M) defined by

$$(73) \quad \begin{aligned} |x_2 + x_3| &\geq a && \text{in } w_M \\ |x_2 + x_3| &< a && \text{elsewhere} \end{aligned}$$

where

$$(74) \quad \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} e^{-\frac{1}{2}x^2} dx = 1 - \alpha$$

$$(75) \quad x_i = \frac{m_i - \frac{1}{4}M}{\frac{1}{2}M^{\frac{1}{2}}} \quad (i = 2, 3),$$

is associated with a test of the hypothesis $H_0(\theta = 0)$ of type A_{∞} at level of significance α .

The proof of this theorem follows the same lines as that of Theorem 1 as far as inequality (42). On putting $P_1 = P_2 = \frac{1}{2}$ in (42) we get

$$(76) \quad \left(-\frac{1}{2}m_2 - \frac{1}{2}m_3 + \frac{1}{4}M - a_3 \right)^2 - \frac{1}{4}(m_2 + m_3 - \frac{1}{2}M) \geq a_4$$

i.e.,

$$(77) \quad (x_2 + x_3 - a_6)^2 \geq a_7.$$

The critical region w_M defined in the statement of the theorem is of this form with $a_6 = 0$ and $a_7 = a^{\frac{1}{2}}$.

Hence the sequence of critical regions (w_M) satisfies conditions (i) of the definition of a test of type A_∞ . The sequence of critical regions may also be shown to satisfy conditions (ii) and (iii) for a test of type A_∞ by following the lines of the proof of THEOREM 1 and noting that $x_2 + x_3 = 2M^{-\frac{1}{2}}(m_2 + m_3 - \frac{1}{2}M)$ tends to be distributed as a unit normal deviate as $M \rightarrow \infty$.

On account of the shape of the critical regions in the general case, I shall for the remainder of this paper call the tests derived in the above theorem the *parabolic tests* for the cases considered.

5. Application of the Parabolic Tests. For practical purposes the formulae derived above are inconvenient to use. I will therefore express them in terms of the deviations of the observed frequencies in the four cells from the frequencies "expected" when the hypothesis $H_0(\theta = 0)$ is true, i.e. in terms of the variables d_i , where

$$(78) \quad d_i = m_i - Mp_{0i} = x_i(Mp_{0i})^{\frac{1}{2}} \quad (i = 1, 2, 3, 4).$$

The test then becomes "reject the hypothesis H_0 at level of significance α if $v + Bu^2 \geq A$ " where

$$(79) \quad u = \frac{d_1(1 - P_1 - P_2) - d_2P_2 - d_3P_1}{\{MP_1P_2(1 - P_1)(1 - P_2)\}^{\frac{1}{2}}}$$

$$(80) \quad v = \frac{P_1(1 - P_1)(2P_2 - 1)(d_1 + d_3) + P_2(1 - P_2)(2P_1 - 1)(d_1 + d_2)}{[MP_1P_2(1 - P_1)(1 - P_2)\{P_1(1 - P_1)(2P_2 - 1)^2 + P_2(1 - P_2)(2P_1 - 1)^2\}]^{\frac{1}{2}}}$$

$$(81) \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha$$

$$(82) \quad B = \left[\frac{MP_1P_2(1 - P_1)(1 - P_2)}{P_1(1 - P_1)(1 - 2P_2)^2 + P_2(1 - P_2)(1 - 2P_1)^2} \right]^{\frac{1}{2}}$$

except when $P_1 = P_2 = \frac{1}{2}$. In the latter case reject the hypothesis H_0 if

$$(83) \quad \left| \frac{d_2 + d_3}{\frac{1}{2}M^{\frac{1}{2}}} \right| \geq a$$

where

$$(84) \quad \frac{1}{\sqrt{2\pi}} \int_a^{+a} e^{-\frac{1}{2}x^2} dx = 1 - \alpha.$$

The application of this last case ($P_1 = P_2 = \frac{1}{2}$) is straightforward. a may be found from the tables of the Normal Probability Integral. d_2 and d_3 may be

calculated from the data, and we may then see whether the inequality (83) is satisfied, and so assess our judgment of the hypothesis H_0 .

TABLE I
Significance of Symbols

A and B are connected by the following relation:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha.$$

Table Ia $\alpha = 0.05$ $p_{.05} = A - 3.8414588 B$		Table Ib $\alpha = 0.01$ $p_{.01} = A - 6.6348966 B$	
B	$p_{.05}$	B	$p_{.01}$
0	1.6449	0	2.3263
1.00	0.322	1.00	0.289
1.25	.256	1.25	.231
1.50	.212	1.50	.192
1.75	.181	1.75	.165
2.00	.158	2.00	.144
2.25	.141	2.25	.128
2.50	.127	2.50	.115
2.75	.116	2.75	.105
3.00	.106	3.00	.096
3.25	.098	3.25	.089
3.50	.091	3.50	.082
3.75	.084	3.75	.077
4.00	.079	4.00	.072
5	.063	5	.058
6	.052	6	.048
7	.045	7	.041
8	.039	8	.036
9	.035	9	.032
10	.031	10	.029
15	.021	15	.020
20	.016	20	.014
30	.010	30	.009
40	.008	40	.007
50	.006	50	.006

The general case is also straightforward, except for the determination of A from equation (81). To facilitate this I have constructed Tables Ia and Ib. These tables correspond respectively to significance levels .05, .01, and from

them the value of A corresponding to a given value of B may be calculated. The quantity tabled, (ρ) , is the difference between A and a multiple¹ (constant for a given level of significance and given with the table to which it applies) of B . To find A , therefore, B is calculated, multiplied by the appropriate constant, and added to the quantity in the table corresponding to B . For large values of B (40 and over) ρ is small, and A may be taken equal to the constant multiple of B .

In particular cases when the values of P_1 and P_2 are substituted in the expression for B (see THEOREM 1 above) and in (79) and (80) above, these equations appear much less formidable. Thus in the case considered by R. A. Fisher [3], $P_1 = P_2 = \frac{1}{4}$ and we get

$$(85) \quad B = \sqrt{\frac{3M}{8}}$$

$$u = \frac{1}{3}M^{-1}(2d_1 - d_2 - d_3); \quad v = -4(6M)^{-1}(2d_1 + d_2 + d_3)$$

and the test becomes "reject the hypothesis H_0 at level of significance α when

$$(86) \quad \phi = \{(2d_1 - d_2 - d_3)^2 - \frac{2}{3}(2d_1 + d_2 + d_3)\} / \{\frac{2}{3}(\frac{3}{8}M)^{\frac{1}{2}}\} \geq A$$

where

$$(87) \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{ e^{-\frac{1}{2}u^2} \int_{A-u^2\sqrt{\frac{1}{3}M}}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du = \alpha.$$

Example. Fisher [3] gives an example of the case $P_1 = P_2 = \frac{1}{4}$. In the series of experiments that he quotes the observed results fall in the four categories respectively as follows:

$$m_1 = 32; \quad m_2 = 904; \quad m_3 = 906; \quad m_4 = 1997. \quad M = 3839.$$

Hence $d_1 = -207.9375$; $d_2 + d_3 = 370.375$. From (86), $\phi = 10863.1$. $B = 37.94239$. From the tables:

$$\text{at } .05 \text{ level, } A_{.05} = 3.8414588 \times 37.94239 + 0.0075 = \underline{145.7615}$$

$$\text{at } .01 \text{ level, } A_{.01} = 6.6348966 \times 37.94239 + 0.0065 = \underline{251.750}.$$

Hence we reject the hypothesis that $\theta = 0$, i.e. that there is no linkage, since the value of ϕ is well outside even the .01 level of significance.

6. Power function of the Tests. General Case. The parabolic test as described above has the desirable property that of all tests (at level of significance α) which are unbiased for large values of M this test will detect small variations in θ most frequently. However, to get a clearer idea of the properties of this

¹ This multiple is equal to $k_{\frac{1}{2}}^2$ where $\frac{1}{\sqrt{2\pi}} \int_{-k_{\frac{1}{2}}}^{+k_{\frac{1}{2}}} e^{-\frac{1}{2}t^2} dt = 1 - \alpha$, α being the level of significance.

test we shall calculate, as accurately as may be practicable, the power function of the test.

As a preliminary step we obtain a rough idea of the power function by making use of the concept of a limiting power function as stated by Neyman [6]. This may be defined as follows:

Let $E_{M'}$ denote the sample point corresponding to a sample of size M' , and put

$$(88) \quad P\{E_{M'} \in w \mid \vartheta'\} = \beta_{M'}(w \mid \vartheta'),$$

where $\vartheta' = M'^{\frac{1}{2}}\theta$, w being a fixed region. Supposing ϑ' kept fixed, let M' increase and let

$$(89) \quad \beta_{\infty}(w \mid \vartheta') = \lim_{M' \rightarrow \infty} \beta_{M'}(w \mid \vartheta')$$

if this limit exists.

Then $\beta_{\infty}(w \mid \vartheta')$ is the limiting power function of the test associated with the critical region w . It will be noted that the limiting power function is a function of ϑ' .

In the problem under consideration the parabolic test when the sample size is M is associated with the critical region w_M . Now it should be noted that in the definition of the limiting power function w remains fixed. Therefore the limiting power function of the parabolic test for sample size M is

$$(90) \quad \beta^{\infty}(w_M \mid \vartheta') = \lim_{M' \rightarrow \infty} \beta_{M'}(w_M \mid \vartheta').$$

The significance of the limiting power function is that for any $\epsilon > 0$ and for any ϑ' there is a number $M_{\epsilon, \vartheta'}$ such that for all $M > M_{\epsilon, \vartheta'}$ we have in our case (by THEOREM 1 of the Appendix)

$$(91) \quad |\beta_M(w_M \mid \vartheta') - \beta_{\infty}(w_M \mid \vartheta')| < \epsilon.$$

It should be noted, however, that the limiting power curve (the graph of the limiting power function against $\theta = \vartheta M^{-\frac{1}{2}}$) may be only a very rough approximation to the actual power curve. Furthermore (Neyman, [6, p. 83]) we cannot, in general, use the limiting power function of a test to answer the question:

"How large must we take our sample size M to detect the falsehood of the hypothesis $H_0(\theta = 0)$ when actually $\theta = \theta'$, with a limiting probability of at least, say, 0.95?"

For if we form a table as below

M	$\vartheta'_{(M)} = M^{\frac{1}{2}}\theta'$	$\beta_{\infty}(w_M \mid \vartheta'_{(M)})$
100
1000
...

it is possible that $\beta_{\infty}(w_M \mid \vartheta'_{(M)})$ may never attain the value 0.95.

THEOREM 3. *The limiting power function of the parabolic test is*

$$(92) \quad \beta_{\infty}(w_M | \vartheta) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}[u - \vartheta(P_1 P_2)^{\frac{1}{2}}(1-P_1)^{-\frac{1}{2}}(1-P_2)^{-\frac{1}{2}}]^2} \int_{A-Bu^2}^{\infty} e^{-\frac{1}{2}v^2} dv \right\} du$$

in all cases for which $0 < P_i < 1$ and P_1 and P_2 are not both equal to $\frac{1}{2}$.

The proof of this theorem follows immediately from THEOREM 1 of the Appendix by applying the transformation (46)–(48) and putting $\lambda = P_1 P_2$.

The above remarks concerning special precautions to be taken with respect to the limiting power function suggest the necessity of studying the actual power function of the parabolic test by some other method.

With this object in view, a study was made of the distribution of the function $\phi = v + Bu^2$ for finite values of M and in particular for $M = 100$ and $M = 3839$. ϕ is a discontinuous variate and, for any given value of M , has definite limits of variation arising from the limitations on the values of the variables m_i stated in the inequalities (25), (26) above. These limits of variation of ϕ were found to be

$$(93) \quad -\frac{4}{3}(\frac{2}{3}M)^{\frac{1}{2}}(\frac{2}{3}M - \frac{1}{18}) < \phi < \frac{4}{3}(\frac{2}{3}M)^{\frac{1}{2}}M(\frac{2}{3}M - 1)$$

for the case $P_1 = P_2 = \frac{1}{4}$. Hence when

$$\begin{aligned} M = 100, \quad & -12.25 < \phi < 5486.86, \\ M = 3839, \quad & -75.89 < \phi < 1310795.75. \end{aligned}$$

Also it was found that

$$(94) \quad \mathfrak{E}(\phi | \theta) = B \left\{ 1 + \frac{(1-2P_1)(1-2P_2)}{(1-P_1)(1-P_2)} (e^{\theta} - 1) + \frac{(M-1)P_1 P_2}{(1-P_1)(1-P_2)} (e^{\theta} - 1)^2 \right\}$$

where $\mathfrak{E}(\phi | \theta)$ denotes the expected value of ϕ , given the value of the parameter θ . Thus when $P_1 = P_2 = \frac{1}{4}$ we have $B = \sqrt{\frac{3}{8}M}$ and so $\mathfrak{E}(\phi | 0) = \sqrt{\frac{3}{8}M}$. Hence when

$$\begin{aligned} M = 100, \quad & \mathfrak{E}(\phi | 0) = 6.12372, \\ M = 3839, \quad & \mathfrak{E}(\phi | 0) = 37.94239. \end{aligned}$$

It is thus seen that the distribution of ϕ might be represented by a Type III curve, since the distribution of ϕ has a finite lower bound and a very long positive tail. In order to fit a Type III curve, we must know the second moment of the curve as well as its lower bound and mean. The general expression for the second moment about zero is too complicated to be printed and so only the numerical expressions obtained by giving special values to M are given below. These are:

(i) $M = 100$

$$(95) \quad \begin{aligned} \mathfrak{E}(\phi^2 | \theta) = & 112.41667 + 165.62963(e^{\theta} - 1) + 2493.33333(e^{\theta} - 1)^2 \\ & + 1078.00000(e^{\theta} - 1)^3 + 4356.91667(e^{\theta} - 1)^4, \end{aligned}$$

(ii) $M = 3839$

$$(96) \quad \xi(\phi^2 | \theta) = 4318.79213 + 6397.29625(e^\theta - 1) + 3684321.24073(e^\theta - 1)^2 \\ + 1636267.33255(e^\theta - 1)^3 + 261530062.11111(e^\theta - 1)^4.$$

Using the above results Type III curves were fitted to the distribution of ϕ , and approximate values of the power functions $\beta(w_M | \theta)$, at level of significance .05, were calculated. This was obtained by evaluating $P\{\phi > A_{.05} | \theta\}$ and assuming the distribution of ϕ to be that given by the fitted curve. Then

$$(97) \quad \beta(w_M | \theta) = P\{\phi > A_{.05} | \theta\}.$$

The values obtained for the limiting and approximate power functions are given in Tables IIa, IIb. Unfortunately the agreement between the two is not satisfactory.

Special Case. For the cases $P_1 = P_2 = \frac{1}{2}$ ($M = 100$, $M = 400$) power functions were calculated on the assumption that for a given value of θ , the random variable $2M^{-1}(d_2 + d_3)$ is distributed normally about a mean $M^{\frac{1}{2}}(e^\theta - 1)$ with standard deviation $\sqrt{e^\theta(2 - e^\theta)}$. This is approximately the case for the values of M considered. The approximate power functions so calculated are given in Tables IIIa, IIIb.

7. Parabolic Test and χ^2 Test. It is interesting to note the close connection between the parabolic test and the χ^2 test as introduced for intuitive reasons and normally used in testing for linkage. The χ^2 test consists of calculating the quantity

$$(98) \quad \chi^2 = \frac{1}{MP_1P_2(1 - P_1)(1 - P_2)} \{(1 - P_1)(1 - P_2)m_1 \\ - P_2(1 - P_1)m_2 - P_1(1 - P_2)m_3 + P_1P_2m_4\}^2$$

and rejecting the hypothesis $H_0(\theta = 0)$ if $|\chi| > a$ where

$$(99) \quad \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} e^{-t^2} dt = 1 - \alpha.$$

In the special case ($P_1 = P_2 = \frac{1}{2}$) the parabolic test and the χ^2 test are identical; while comparing (98) and (79) we see that in the general case

$$(100) \quad u = \chi.$$

Hence in the general case the criterion used in the parabolic test may be written

$$(101) \quad \phi = v + B\chi^2.$$

(1) *Large Samples.* For large samples the first term of the expression $v + B\chi^2$ is usually of small importance, since

v is of form $M^{-1} \times$ (linear function of the d_i 's), while $B\chi^2$ is of form $M^{-1} \times$ (quadratic function of the d_i 's).

For such samples the χ^2 test and parabolic test would appear to be nearly equivalent.

TABLE II
Limiting and Approximate Power Functions of Parabolic Test

$$P_1 = P_2 = \frac{1}{4} \\ -\infty < \theta < 1.386$$

Table IIa
 $M = 100$

θ	Power	
	Limiting	Approximate
-2.00		0.90870
-1.50	0.99880	
-1.40		0.77656
-1.20	0.97915	0.69505
-1.05	0.93786	
-1.00		0.58580
-0.90	0.85024	
-0.75	0.70467	0.42755
-0.60	0.51532	
-0.45	0.32258	0.21849
-0.30	0.16986	0.12504
-0.15	0.07905	0.05689
-0.10	0.06280	0.04438
-0.05	0.05318	0.03866
0.00	0.05000	0.04069
0.05	0.05318	0.05021
0.10	0.06280	0.07429
0.15	0.07905	
0.30	0.16986	0.26559
0.45	0.32258	
0.60	0.51532	0.75854
0.75	0.70467	0.94245

Table IIb
 $M = 3839$

θ	Power	
	Limiting	Approximate
-0.25	0.99932	0.99853
-0.20	0.98502	0.97521
-0.15	0.87243	0.83620
-0.10	0.54197	0.52066
-0.05	0.17827	0.19223
0.00	0.05000	0.04111
0.05	0.17827	0.21568
0.10	0.54197	0.59517
0.15	0.87243	0.91641
0.20	0.98502	0.99640
0.25	0.99932	0.99999

THEOREM 4. *The limiting power function of the χ^2 test is*

$$(102) \quad \beta_\infty(w_\chi^2 | \vartheta) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} e^{-\frac{1}{2}(u - \vartheta(P_1 P_2)^{\frac{1}{2}}(1-P_1)^{-\frac{1}{2}}(1-P_2)^{-\frac{1}{2}})^2} du$$

(w_χ^2 denotes the region defined by the inequality $|\chi| > a$).

This theorem may be proved by applying (46)–(48) to $Q_\vartheta(x_1, x_2, x_3)$ in THEOREM 1 of the Appendix, and noting that $u = \chi$ by (100).

We notice that $\beta_{\infty}(w_{\chi^2} | \vartheta)$, for a given value of ϑ , has the same value for all values of M , unlike the limiting power function $\beta_{\infty}(w_M | \vartheta)$ of the parabolic test. It is this point which accounts for the seeming paradox that, despite the manner in which the parabolic test was defined, for all values of ϑ and M

$$(103) \quad \beta_{\infty}(w_{\chi^2} | \vartheta) \geq \beta_{\infty}(w_M | \vartheta)$$

as may be deduced from (92) and (102). This does not mean that for any given ϑ and all M sufficiently large the power function of the χ^2 test, $\beta_M(w_{\chi^2} | \vartheta)$,

TABLE III
Approximate Power Function

$$P_1 = P_2 = \frac{1}{2}$$

$$-\infty < \theta < 0.693$$

<i>Table IIIa.</i> $M = 100$		<i>Table IIIb.</i> $M = 400$	
θ	Power	θ	Power
-0.45	0.96288	-0.25	0.99424
-0.40	0.92161	-0.20	0.95482
-0.35	0.85072	-0.15	0.79787
-0.30	0.74351	-0.10	0.47734
-0.25	0.60197	-0.05	0.16378
-0.20	0.44054	-0.02	0.06810
-0.15	0.28380	0.00	0.05000
-0.10	0.15727	0.02	0.06885
-0.05	0.07737	0.05	0.17609
0.00	0.05000	0.10	0.55737
0.05	0.08029	0.15	0.90213
0.10	0.18177	0.20	0.99431
0.15	0.36464	0.25	0.99995
0.20	0.60278		
0.25	0.82071		
0.30	0.94975		
0.35	0.99299		

is necessarily not less than the power function of the parabolic test, $\beta_M(w_M | \vartheta)$. For although, given any $\epsilon > 0$, there is a number $M_{\epsilon, \vartheta}$ such that if $M > M_{\epsilon, \vartheta}$

$$(104) \quad |\beta_M(w_{\chi^2} | \vartheta) - \beta_{\infty}(w_{\chi^2} | \vartheta)| < \epsilon$$

and

$$(105) \quad |\beta_M(w_M | \vartheta) - \beta_{\infty}(w_M | \vartheta)| < \epsilon$$

it may be that for such values of $M_{\epsilon, \vartheta}$

$$(106) \quad 0 \leq \beta_{\infty}(w_{\chi^2} | \vartheta) - \beta_{\infty}(w_M | \vartheta) < 2\epsilon.$$

The above results show, however, how close the agreement between the power functions of the two tests is for large values of M . In fact we have

$$(107) \quad \lim_{M \rightarrow \infty} \beta_{\infty}(w_M | \vartheta) = \beta_{\infty}(w_{\chi^2} | \vartheta).$$

This may be easily proved, since as M increases w_M approximates to w_{χ^2} .

(2) *Small Samples.* In order to obtain some idea of the relations between the two tests when M is small (i.e. less than 100), the case $P_1 = P_2 = \frac{1}{4}$, $M = 32$ was considered in some detail.

In this case our tests at 5% level of significance are respectively χ^2 test, reject if

$$(108) \quad |2y - z| > 8.315$$

parabolic test, reject if

$$(109) \quad (2y - z)^2 - \frac{3}{2}(2y + z) > 69.576$$

where

$$(110) \quad y = d_1 \quad z = d_2 + d_3.$$

All samples for which the verdicts of the two above tests would not agree were obtained. These were as follows:

(a) Samples for which H_0 is accepted by χ^2 test, rejected by parabolic test

$y =$	1	0	-1	-2	Probability of drawing sample of this type when H_0 is true is 0.00320.
$z =$	-6	-8	-10	-12	

(b) Samples for which H_0 is rejected by parabolic test, accepted by χ^2 test

$y =$	0	1	2	3	5	6	7	8	8	9	9	Probability of drawing sample of this type when H_0 is true is 0.00038.
$z =$	9	11	13	15	1	3	5	6	7	8	9	

Thus the probability of the two tests giving different verdicts when H_0 is in fact true is only 0.00358.

It will be noted that the above results imply that

$$(111) \quad \beta_{32}(w_{32} | 0) - \beta_{32}(w_{\chi^2} | 0) = 0.00320 - 0.00038 = 0.00282;$$

i.e. that the true levels of significance of the two tests are not equal. This is to be expected, because of the discontinuity of the probability distribution of sample points, which makes it unlikely that the level of significance of either test is exactly .05.

Similarly we can obtain values of $\beta_{32}(w_{32} | \theta) - \beta_{32}(w_{\chi^2} | \theta)$, the differences in the powers of the two tests with respect to various alternative hypotheses. These values were obtained for a few values of θ .

θ	$\beta_{32}(w_{32} \theta) - \beta_{32}(w_{\chi^2} \theta)$
-0.5	0.01625
0.0	0.00282
0.5	-0.00006

These figures indicate that the parabolic test detects negative θ 's better than the χ^2 test, but that the χ^2 test detects positive θ 's better than the parabolic test, although the advantage in this latter case is minute.

The critical regions associated with the two tests may be represented by regions in the (y, z) plane. The critical region for the parabolic test will be defined by

$$(112) \quad (2y - z)^2 - \frac{3}{2}(2y + z) > \nu$$

and that for the χ^2 test, w_{χ^2} , by

$$(113) \quad (2y - z)^2 > \nu'$$

where $\nu \approx \nu'$.

w_{χ^2} is therefore the complement of the region lying between the lines L_1, L_2 with equations $2y - z = \pm\sqrt{\nu'}$; w_M lies outside the parabola K with equation $(2y - z)^2 - \frac{3}{2}(2y + z) = \nu$.

Since $\nu \approx \nu'$, K meets L_1, L_2 at points near the respective intersections of L_1, L_2 with the line $2y + z = 0$. See Figure 1.

In the diagram the regions V_1, V_2 contain all sample points for which the χ^2 test rejects and the parabolic test accepts H_0 ; U_1, U_2 contain all sample points for which the χ^2 test accepts and the parabolic test rejects H_0 .

For a given value of θ it is known that the probability distribution is approximately such that the quantity

$$(114) \quad \psi_\theta^2 = \frac{\{y - \frac{1}{16}M(e^\theta - 1)\}^2}{\frac{1}{16}M + \frac{1}{16}M(e^\theta - 1)} + \frac{\{z + \frac{1}{8}M(e^\theta - 1)\}^2}{\frac{1}{16}M - \frac{1}{8}M(e^\theta - 1)} + \frac{\{y + z + \frac{1}{16}M(e^\theta - 1)\}^2}{\frac{1}{16}M + \frac{1}{16}M(e^\theta - 1)}$$

is distributed as χ^2 with 2 degrees of freedom.

The ellipses of equal density $\psi_\theta^2 = \text{constant}$ have centers at points $(\frac{1}{16}M[e^\theta - 1], -\frac{1}{8}M[e^\theta - 1])$ which must lie on the line $2y + z = 0$. When $\theta = 0$ the center is at the origin, and the major and minor axes of the ellipse make angles of approximately 99.5° and 9.5° respectively with the y -axis. For small changes in θ the angles of inclination of the major and minor axes of the ellipse to the coordinate axes are not greatly changed, and we see that as the center of the ellipse moves along the line $2y + z = 0$ we have

(1) θ increasing: center moves downwards, tending to increase $P\{E \in U_2\} - P\{E \in V_2\}$ while $P\{E \in V_1\}$ and $P\{E \in U_1\}$ both become small. Thus $\beta_M(w_M | \theta)$ tends to increase quicker than $\beta_M(w_{\chi^2} | \theta)$.

(2) θ decreasing: here we have the opposite effect and $\beta_M(w_M | \theta)$ tends to increase slower than $\beta_M(w_{X^2} | \theta)$.

These conclusions agree qualitatively with those drawn in the case $M = 32$. (N.B. In the case $M = 32$ no sample points fall into the region U_1 because no points in U_1 satisfy the inequalities (25), (26)).

8. Some Geometrical Considerations. In this section we shall consider the manner in which the situations dealt with above may be interpreted in terms

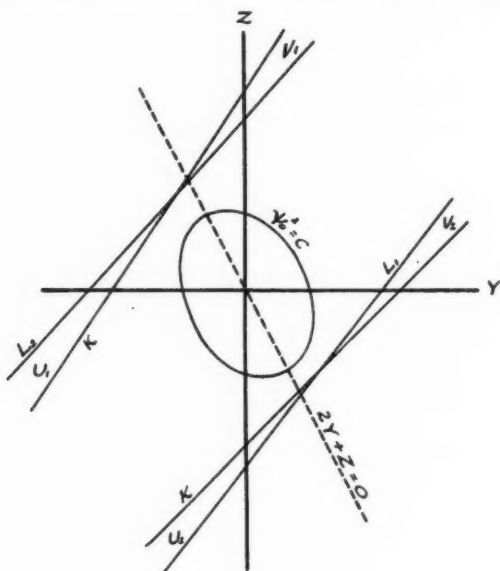


FIG. 1

of geometrical concepts. It will be convenient to consider as variables $n_i = m_i/M$. The sample space $W(n)$ is then bounded by the four planes

$$(115) \quad \begin{aligned} n_i &= 0 & (i = 1, 2, 3), \\ \sum_{i=1}^3 n_i &= 1. \end{aligned}$$

In this space, corresponding to any admissible hypothesis H_θ specifying a value of θ , there is a point T_θ with coordinates $(\theta^{n_1}, \theta^{n_2}, \theta^{n_3})$ where

$$(116) \quad \begin{aligned} \theta^{n_1} &= P_1 P_2 e^\theta, \\ \theta^{n_2} &= P_1 (1 - P_2 e^\theta), \\ \theta^{n_3} &= P_2 (1 - P_1 e^\theta). \end{aligned}$$

These are the proportions of results expected in the first three cells, if the hypothesis H_θ specifying θ be true.

Now, if H_θ be true, we have

$$(117) \quad P\{n_1 = n'_1, n_2 = n'_2, n_3 = n'_3, n_4 = n'_4 \mid H_\theta\} \approx ce^{-\lambda x^2}$$

where c is constant for a fixed sample size M , and

$$(118) \quad \frac{\chi^2_\theta}{M} = \sum_{i=1}^3 \frac{(n'_i - \theta^{n_i})^2}{\theta^{n_i}} + \frac{\left[\sum_{i=1}^3 (n'_i - \theta^{n_i}) \right]^2}{1 - \sum_{i=1}^3 \theta^{n_i}}.$$

Hence the most frequent position(s) of the sample point E will be somewhere near the point T_θ , which I shall therefore call the *center of density*. It will be noticed that, whatever be the value of θ , the point T_θ must lie on the line

$$(119) \quad n_1 - P_1 P_2 = -[n_2 - P_1(1 - P_2)] = -[n_3 - P_2(1 - P_1)].$$

This line, a segment of which is the locus of the center of density for our set of admissible hypotheses, will be called the *line of density*.

In this space the parabolic test corresponds to a critical region comprising the exterior of a parabolic cylinder. The equation of the boundary of this critical region at level of significance .05 was found for the case $P_1 = P_2 = \frac{1}{4}$, and a model made of it. Also included in the model were the ellipsoids

$$(120) \quad \chi^2_\theta = K_{.05}$$

where $K_{.05}$ is a constant so chosen that

$$(121) \quad P\{\chi^2_\theta > K_{.05} \mid \theta\} \approx .05$$

corresponding to

(i) the case when H_0 is true

(ii) the cases when

$$(122) \quad (a) \quad p_1 = \frac{3}{32}; p_2 = p_3 = \frac{5}{32}; p_4 = \frac{19}{32} \quad \text{i.e. } \theta = 0.41$$

$$(123) \quad (b) \quad p_1 = \frac{1}{32}; p_2 = p_3 = \frac{7}{32}; p_4 = \frac{17}{32} \quad \text{i.e. } \theta = -0.69.$$

It was found that in the case $P_1 = P_2 = \frac{1}{4}$ one axis of all the χ^2_θ -ellipsoids was perpendicular to the plane through the line of density and the axis of n_i . The generators of the boundary of the parabolic acceptance region are also perpendicular to this plane. (By "acceptance region" is meant the complement of the critical region. The acceptance region may be written symbolically \bar{w}_M .) There were further added to the model the intersections with this plane of the ellipsoids at probability level .01, corresponding to the three hypotheses considered above ($\theta = 0, 0.41, -0.69$) and two others, viz.

$$(124) \quad p_1 = \frac{5}{32}; p_2 = p_3 = \frac{3}{32}; p_4 = \frac{21}{32} \quad \text{i.e. } \theta = 0.92,$$

$$(125) \quad p_1 = \frac{1}{64}; p_2 = p_3 = \frac{15}{64}; p_4 = \frac{33}{64} \quad \text{i.e. } \theta = -1.39.$$

For convenience in making the model to a simple scale (1 unit \equiv 150 cms.) it was found necessary to take the sample size M as 1312.5. The model is shown in Figure 2. It will be seen that the acceptance region for the parabolic test is approximately enclosed between two parallel planes perpendicular to the plane common to the line of density and the axis of n_1 . These two planes, in fact, enclose the acceptance region for the χ^2 test. The vertex of the normal

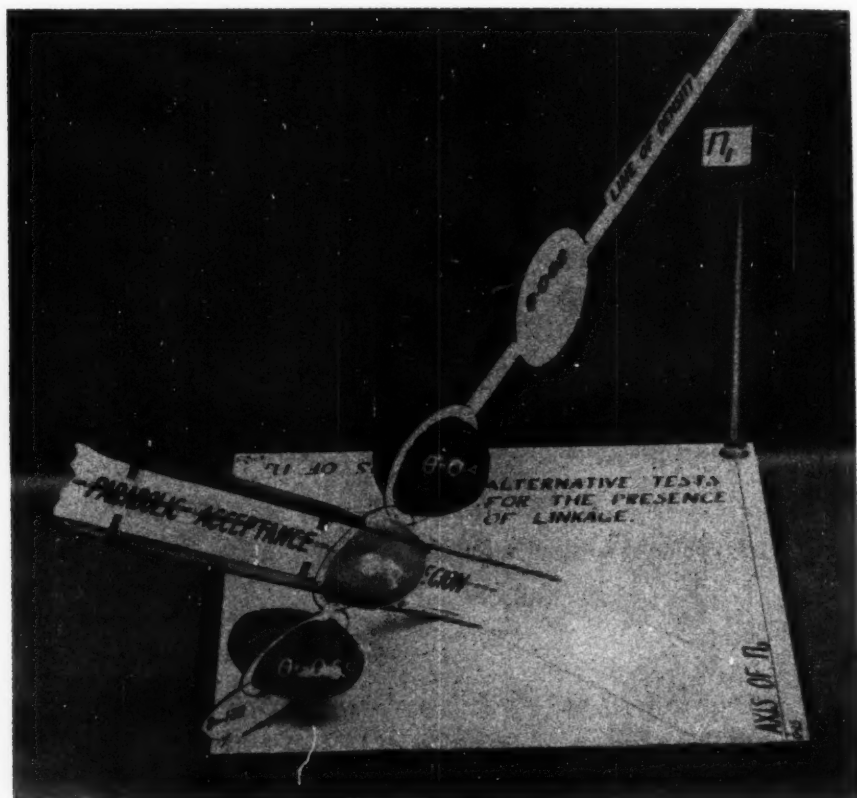


FIG. 2

parabolic section of the parabolic acceptance region is at a comparatively great distance "below" the plane $n_1 = 0$.

As an interesting digression we may use our model to compare qualitatively the parabolic test with yet a third possible test of H_0 . This test is to reject H_0 at level of significance .05 if

$$(126) \quad \chi_0^2 > K_{.05}$$

and may be called the χ_0^2 test. The χ_0^2 -ellipsoid shown in the model is the acceptance region for this test. It will be noticed that when $\theta \neq 0$ the ellipsoids

of equal density include somewhat more of the acceptance region of the χ_0^2 test than of the parabolic acceptance region. This means that the χ_0^2 test would detect that the hypothesis $H_0(\theta = 0)$ is false in these cases, less frequently than would the parabolic and χ^2 tests. We also notice that the center of density T_θ leaves the parabolic acceptance region before it leaves the acceptance region of the χ_0^2 test as it moves along the line of density from the point where $\theta = 0$, whether the direction of motion of T_θ corresponds to θ increasing or decreasing. This also indicates that the χ_0^2 test would act less efficiently than the other two tests.

9. Appendix. In this appendix are obtained various results which, while essential to the main argument, would appear as digressions if they were interpolated as required. The numbering of equations in this appendix does not continue from that of the previous sections, but forms a separate group.

LEMMA. If $f_0(m), f_1(m), \dots, f_s(m)$ be $(s + 1)$ functions of the k variables m_1, m_2, \dots, m_k which are zero except for a finite number of sets of integral values of m_1, \dots, m_k ; and if w_0 be a region in the space of m 's such that

$$(1) \quad f_0(m) \geq \sum_{i=1}^s a_i f_i(m) \quad \text{in } w_0$$

$$(2) \quad f_0(m) < \sum_{i=1}^s a_i f_i(m) \quad \text{in } \bar{w}_0$$

a_1, a_2, \dots, a_s being arbitrary constants; then if w be any region such that

$$(3) \quad \sum_w f_i(m) = \sum_{w_0} f_i(m) \quad (i = 1, \dots, s),$$

we shall have

$$(4) \quad \sum_w f_0(m) \leq \sum_{w_0} f_0(m).$$

PROOF. Let

$$(5) \quad \begin{aligned} \delta &= \sum_{w_0} f_0(m) - \sum_w f_0(m) \\ &= \sum_{w_0 - ww_0} f_0(m) - \sum_{w - ww_0} f_0(m) \end{aligned}$$

where ww_0 denotes the common part of w and w_0 .

Hence the region $w - ww_0$, consisting of those points of w which are not in ww_0 , and so not in w_0 , is contained in \bar{w}_0 . Similarly the region $w_0 - ww_0$ is contained in w_0 . Hence, by inequalities (1),

$$(6) \quad \delta \geq \sum_{w_0 - ww_0} \left\{ \sum_{i=1}^s a_i f_i(m) \right\} - \sum_{w - ww_0} \left\{ \sum_{i=1}^s a_i f_i(m) \right\}$$

and so

$$(7) \quad \delta \geq \sum_{w_0} \left\{ \sum_{i=1}^s a_i f_i(m) \right\} - \sum_w \left\{ \sum_{i=1}^s a_i f_i(m) \right\}.$$

Since the total number of terms in each double summation is finite, we have

$$(8) \quad \delta \geq \sum_{i=1}^s a_i \left\{ \sum_{w_0} f_i(m)' - \sum_w f_i(m) \right\}.$$

But

$$(9) \quad \sum_{w_0} f_i(m) = \sum_w f_i(m), \quad (i = 1, \dots, s).$$

Hence

$$\delta \geq 0, \quad \text{and} \quad \sum_w f_0(m) \leq \sum_{w_0} f_0(m).$$

A lemma similar to the lemma above, where the f 's are taken to be integrable functions and summation over the regions w , w_0 is replaced by integration over these regions, is given by Neyman and Pearson [9]. The proof given above follows the lines of the proof given in that paper.

THEOREM 1. *Suppose that, in a quadrinomial population:*

(i) *the cell probabilities are dependent on the number M of trials made, and are given by*

$$(10) \quad \begin{aligned} p_1 &= p_{01} + \varphi_M \\ p_2 &= p_{02} - \varphi_M \\ p_3 &= p_{03} - \varphi_M \\ p_4 &= p_{04} + \varphi_M \end{aligned}$$

where

$$(11) \quad \sum_{i=1}^4 p_{0i} = \sum_{i=1}^4 p_i = 1$$

and

$$(12) \quad \varphi_M = \lambda(e^{\lambda M^{-1}} - 1)$$

(ii)

$$(13) \quad x_i = (m_i - M p_{0i}) / (M p_{0i})^{\frac{1}{2}} \quad (i = 1, 2, 3, 4)$$

where m_i = number of results falling in i -th cell.

(iii) $w(x)$, or briefly w , is a region in the space W of x_1, x_2, x_3 ; and $P_M(w)$ is the integral probability law of w corresponding to the values p_1, p_2, p_3, p_4 of the cell probabilities given in (2) above when we have M independent trials.

Then

$$(14) \quad P_M(w) \rightarrow \frac{1}{(2\pi)^{\frac{3}{2}} p_{04}^{\frac{1}{2}}} \iiint_w e^{-\frac{1}{2} Q_0(x_1, x_2, x_3)} dx_1 dx_2 dx_3$$

uniformly over W as $M \rightarrow \infty$, where

$$\begin{aligned}
 Q_{\theta}(x_1, x_2, x_3) = & \sum_{i=1}^3 x_i^2(1 + p_{0i} p_{04}^{-1}) + 2p_{04}^{-1} \sum_{i < j \leq 3} x_i x_j (p_{0i} p_{0j})^{\frac{1}{2}} \\
 (15) \quad & - 2\lambda \theta \{x_1(p_{01}^{-1} - p_{02} p_{04}^{-1}) - x_2(p_{02}^{-1} + p_{02} p_{04}^{-1}) \\
 & - x_3(p_{03}^{-1} + p_{03} p_{04}^{-1})\} + \lambda^2 \theta^2 \sum_{i=1}^4 p_{0i}^{-1}.
 \end{aligned}$$

This theorem may be proved by the same method as that used by F. N. David [2] in proving the generalized theorem of Laplace.

I would like to thank Professor Neyman for his invaluable suggestions and advice in the preparation of this paper.

REFERENCES

- [1] JOSEPH F. DALY, *Annals of Math. Stat.*, Vol. 11 (1940), p. 1.
- [2] F. N. DAVID, *Stat. Res. Mem.*, Vol. 2 (1938), p. 69.
- [3] R. A. FISHER, *Statistical Methods for Research Workers*, 6th. ed. (1936), Oliver and Boyd, London.
- [4] J. NEYMAN, *Phil. Trans.*, A., Vol. 236 (1937), p. 333.
- [5] J. NEYMAN, *Skand. Actuar. Tidskr.*, Vol. 20 (1937), p. 149.
- [6] J. NEYMAN, *Ann. Math. Stat.*, Vol. 9 (1938), p. 69.
- [7] J. NEYMAN AND E. S. PEARSON, *Biometrika*, Vol. 20A (1928), p. 223.
- [8] J. NEYMAN AND E. S. PEARSON, *Phil. Trans.*, A, Vol. 231 (1933), p. 289.
- [9] J. NEYMAN AND E. S. PEARSON, *Stat. Res. Mem.*, Vol. 1 (1936), p. 1.
- [10] J. NEYMAN AND E. S. PEARSON, *Stat. Res. Mem.*, Vol. 2 (1938), p. 25.

DEPARTMENT OF STATISTICS,
UNIVERSITY COLLEGE,
LONDON, ENGLAND

REDUCTION OF A CERTAIN CLASS OF COMPOSITE STATISTICAL HYPOTHESES

BY GEORGE W. BROWN

1. Introduction. A situation frequently met in sampling theory is the following: x has distribution $f(x, \theta)$, where θ is an unknown parameter, and for samples (x_1, \dots, x_n) there exists in the sample space E_n a family of $(n - 1)$ -dimensional manifolds upon each of which the distribution is independent of θ ; in addition there is a residual one-dimensional manifold available for estimating θ . For example, suppose there exists a sufficient statistic T for θ , then on the manifolds $T = T_0$ there is defined an induced distribution which is independent of the parameter.

A similar situation is observed when θ is a "location" or "scale" parameter. Let x have the distribution $f(x - a)$ for some a , then the set $(x_2 - x_1, x_3 - x_1, \dots, x_n - x_1)$, or any equivalent set, such as $(x_2 - \bar{x}, \dots, x_n - \bar{x})$, have a joint distribution independent of a , and there is a residual distribution corresponding to each particular configuration $(x_2 - x_1, \dots, x_n - x_1)$. Fisher [1] and Pitman [5] have examined the residual distributions in connection with the problem of estimating scale and location parameters. In this paper we shall be concerned primarily, not with the residual distribution, but with the remainder of the sample information, corresponding to the $(n - 1)$ -dimensional distribution which is independent of the parameter. It is found, in a rather broad class of distributions, that the part of the sample *not* used for estimation *determines, except for the parameter value, the original functional form of the distribution of x .*

This paper is devoted mainly to a study of particular classes of distributions having the property mentioned above. We consider also the theoretical application of this property to certain types of *composite* hypotheses which may be reduced thereby to equivalent *simple* hypotheses.¹ The principal results of this nature may be summed up as follows: If x has distribution of the form $f(x, \theta)$, where θ is either a location or scale parameter, or a vector denoting both, then there exists, in samples (x_1, \dots, x_n) a set of functions $y_i(x_1, \dots, x_n)$, $i = 1, 2, \dots, p$, $p < n$, having joint distribution $D(y_1, \dots, y_p)$ independent of θ , and such that the converse statement holds, namely, *if $\{y_i\}$ have the distribution $D(y_1, \dots, y_p)$, then x has, for some θ , a distribution of the form $f(x, \theta)$.* There is a corresponding statement when x has a distribution of the form $f(x - \sum a_i u_i)$, where the $\{a_i\}$ are parameters, and the $\{u_i\}$ are regression variables.

¹ We use the terms simple and composite hypotheses in the sense of Neyman and Pearson [2].

2. Location and Scale. This section is devoted to the study of functions of the sample observations which are such that their distributions determine the distribution of x , except possibly for location and scale.

It will be assumed that associated with x there is a function $F(x)$ such that

(a) $F(x)$ is monotone non-decreasing,

(b) $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$, and (c) $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$

with the normalization $F(x)$ upper semi-continuous. $F(x)$ is the probability that the random variate takes a value less than or equal to x . If $F(x)$ is associated with the random variate x we say that x has the distribution $F(x)$.

If $g(x)$ is a Borel-measurable function, the Lebesgue-Stieltjes integral $\int_{-\infty}^{\infty} g(x) dF(x)$ is denoted by $E[g(x)]$.

The characteristic function $\varphi(t) = E(e^{itx})$ determines $F(x)$, that is, if $\int_{-\infty}^{\infty} e^{itz} dG(x) = \int_{-\infty}^{\infty} e^{itz} dF(x)$, then $F(x) = G(x)$.

Similarly, let $F(x_1, \dots, x_k)$ be such that

(a) $F(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_k) \geq F(x_1, \dots, x_i, \dots, x_k)$ for $h > 0$ and $i = 1, 2, \dots, k$;

(b) $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_k) = 0, i = 1, 2, \dots, k$;

(c) $\lim_{x_1, \dots, x_k \rightarrow \infty} F(x_1, \dots, x_k) = 1$;

with the normalization $F(x_1, \dots, x_k)$ continuous on the right in each x_i . If $F(x_1, \dots, x_k)$ is associated with x_1, \dots, x_k we say that x_1, \dots, x_k have the

joint distribution $F(x_1, \dots, x_k)$. As before, $E[H(x_1, \dots, x_k)] = \int_{R_k} H dF$,

where R_k is the Euclidean k -space. It is well known that under such conditions, given Borel-measurable functions $y_i(x_1, \dots, x_k), i = 1, \dots, p, p \leq k$,

then $G(y_1, \dots, y_p) = \int_{R(y)} dF(x_1, \dots, x_k)$, where $R(y)$ is the region $[y_1(x_1, \dots, x_k) \leq y_1, \dots, y_p(x_1, \dots, x_k) \leq y_p]$, is again a distribution function satisfying the conditions above. Moreover, $\int_R g(y_1, \dots, y_p) dG(y_1, \dots, y_p) =$

$\int_{R'} g[y_1(x_1, \dots, x_k), \dots, y_p(x_1, \dots, x_k)] dF$, where R' is the set of all points (x_1, \dots, x_k) such that $[y_1(x_1, \dots, x_k), \dots, y_p(x_1, \dots, x_k)] \in R$.

If x has distribution $F(x)$, then, by definition, the set (x_1, \dots, x_n) is a sample from this distribution if x_1, \dots, x_n have the joint distribution $F(x_1) \dots F(x_n)$.

The following theorem states that two distributions giving rise, in sampling, to the same distribution of the set $x_1 - x_n, x_2 - x_n, \dots, x_{n-1} - x_n$, with $n \geq 3$, can differ at most by a translation, that is, the distribution of that set determines the original distribution except for location.

THEOREM 1A: Let x have the distribution $F(x)$. Denote by S the set of zeros of

$\int e^{itz} dF(x)$ and denote by ϵ the g.l.b. of $|t|$ for t in S . Suppose that the complement of S is ϵ -connected.² Suppose that x' has distribution $G(x')$, and let x_1, \dots, x_n and x'_1, \dots, x'_n be samples. Then the set $w_\alpha = x_\alpha - x_n$, $\alpha = 1, \dots, n-1$, have the same joint distribution as the set $w'_\alpha = x'_\alpha - x'_n$ if and only if there exists a constant a such that $x' + a$ and x have the same distribution.

PROOF: The sufficiency of the condition follows immediately, since $w'_\alpha = x'_\alpha - x'_n = (x'_\alpha + a) - (x'_n + a)$.

In establishing necessity, only the fact that w_1, w_2 have the same joint distribution as w'_1, w'_2 is needed. This hypothesis implies that

$$E\{e^{it_1 w_1 + it_2 w_2}\} = E\{e^{it_1 w'_1 + it_2 w'_2}\},$$

that is,

$$E\{e^{it_1(x_1 - x_n) + it_2(x_2 - x_n)}\} = E\{e^{it_1(x'_1 - x'_n) + it_2(x'_2 - x'_n)}\}.$$

Set $\varphi(t) = E(e^{itz})$, $\psi(t) = E(e^{itx'})$. The relation above becomes

$$(1) \quad \varphi(t_1)\varphi(t_2)\varphi(-t_1 - t_2) = \psi(t_1)\psi(t_2)\psi(-t_1 - t_2).$$

Consider equation (1) for values of t_1, t_2 in the neighborhood of $t = 0$. $\varphi(0) = \psi(0) = 1$, hence there is an interval $|t| < \delta$, in which $\varphi(t)$ and $\psi(t)$ do not vanish. It is easily shown that $\varphi(t)$ and $\psi(t)$ are each continuous, since e^{itz} , in the neighborhood of $t = 0$, is continuous uniformly for any bounded interval of x , and since A may be chosen so that $1 - F(A)$ and $F(-A)$ are both as small as desired. In the interval $|t| < \delta$ the function $f(t) = \varphi(t)/\psi(t)$ is continuous. Also, $\varphi(-t) = \overline{\varphi(t)}$ and $\psi(-t) = \overline{\psi(t)}$. Setting $t_2 = 0$ in (1) we obtain $\varphi(t)\varphi(-t) = \psi(t)\psi(-t)$, hence $|\varphi(t)| = |\psi(t)|$, that is, $|f(t)| = 1$. $f(t)$ takes values on the unit circle of the complex plane, and $f(0) = 1$, hence there is an interval $|t| < \delta'$ such that $z = f(t)$ lies on an arc γ , of length less than 2π , containing the point $z = 1$. Now consider the functional equation (1) for $|t_1| < \frac{1}{2}\delta'$, $|t_2| < \frac{1}{2}\delta'$. (1) becomes

$$f(t_1)f(t_2)f(-t_1 - t_2) = 1.$$

The interval $|t| < \delta'$ was so chosen that for $|t_1| < \frac{1}{2}\delta'$, $|t_2| < \frac{1}{2}\delta'$, it is possible to define a single-valued branch of the argument of $f(t_1)$, $f(t_2)$, and $f(t_1 + t_2)$. Letting $t_2 = 0$ we have $f(t)f(-t) = 1$, hence, replacing $f(-t_1 - t_2)$ by $1/f(t_1 + t_2)$ in the last equation, we have

$$f(t_1)f(t_2) = f(t_1 + t_2).$$

Arg $f(t_1)$, arg $f(t_2)$, and arg $f(t_1 + t_2)$ are uniquely determined, except for some fixed multiple of 2π . If we choose the principal value of the argument, i.e., so

² The set S is ϵ -connected if any two points p, q , in S can be connected by an ϵ -chain, i.e., there exists a set $p_0 = p, p_1, \dots, p_{n-1}, p_n = q$, such that $|p_i - p_{i-1}| < \epsilon$, $i = 1, 2, \dots, n$.

that $0 \leq \arg f(t) < 2\pi$, we must have

$$\arg f(t_1) + \arg f(t_2) = \arg f(t_1 + t_2)$$

for $|t_1| < \frac{1}{2}\delta'$, $|t_2| < \frac{1}{2}\delta'$. Since $\arg f(t)$ is continuous, any solution of this well known functional equation must be of the form $\arg f(t) = at$. $|f(t)| = 1$, therefore there exists a constant a such that $f(t) = e^{iat}$, for $|t| < \frac{1}{2}\delta'$, that is, $\varphi(t) = e^{iat}\psi(t)$, for $|t| < \frac{1}{2}\delta'$. By use of (1) this may be extended to hold for all t such that $|t| < \epsilon$, where ϵ is the minimum modulus of all t such that $\varphi(t) = 0$. (1) may now be used to extend the relation for all t such that $\varphi(t) \neq 0$ by choosing an ϵ -chain connecting the origin to the point t . We know already that $\varphi(t) = e^{iat}\psi(t)$ if $\varphi(t) \neq 0$, hence it holds for all t . This relation says that $E(e^{itx}) = E(e^{it(x'+a)})$, hence $x' + a$ and x have the same distribution, thus completing the demonstration of the theorem.

It should be remarked that the set $(x_1 - x_n, \dots, x_{n-1} - x_n)$ may be replaced in Theorem Ia by any equivalent set, for example, $(x_1 - \bar{x}, \dots, x_{n-1} - \bar{x})$.

The next result is of the same nature as Theorem Ia except for the replacement of the location parameter by a scale (positive or negative) parameter.

THEOREM IB: Let x have distribution $F(x)$, such that the zeros of $\int_{-\infty}^{\infty} e^{it(\log|x|)} dF(x)$ are nowhere dense, and let x' have distribution $G(x')$. Let x_1, \dots, x_n and x'_1, \dots, x'_n be samples from the distributions of x and x' , with $n \geq 3$, then the set $w_\alpha = x_\alpha/x_n$, $\alpha = 1, \dots, n-1$, have the same distribution as the set $w'_\alpha = x'_\alpha/x'_n$ if and only if there exists a constant c such that cx' and x have the same distribution.

PROOF: The sufficiency of the condition is evident. Suppose, then, as before, that w_1, w_2 have the same joint distribution as w'_1, w'_2 . $\log|w_1|$ and $\log|w_2|$ have the same joint distribution as $\log|w'_1|$ and $\log|w'_2|$, hence by application of Theorem Ia to $\log|x|$ and $\log|x'|$ it follows (since the complement of a nowhere dense set is ϵ -connected for every ϵ) that there exists a constant a such that

$$\int_{-\infty}^{\infty} e^{it \log|x|} dF(x) = \int_{-\infty}^{\infty} e^{it(\log|x'| - a)} dG(x').$$

Let $y = e^{-a}x'$, then $|x|$ and $|y|$ have the same distribution, and

$$(2) \quad \int e^{it \log|x|} dF(x) = \int e^{it \log|y|} dH(y),$$

where y has distribution $H(y)$. We now have to show that either y or $-y$ has the distribution of x , that is, it must be shown that either $H(y) = F(y)$, or $H(y) = 1 - F(-y)$.

By the first part of the theorem the functions $u_1 = y_1/y_3$ and $u_2 = y_2/y_3$ have the same joint distribution as w_1, w_2 . It is clear that the mean value of any function of u_1 and u_2 is the same as the mean value of the corresponding func-

tion of w_1 and w_2 . Hence

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i[t_1 \log |w_1| + t_2 \log |w_2|]} \operatorname{sgn} w_1 \operatorname{sgn} w_2 dF(x_1) dF(x_2) dF(x_3) \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i[t_1 \log |u_1| + t_2 \log |u_2|]} \operatorname{sgn} u_1 \operatorname{sgn} u_2 dH(y_1) dH(y_2) dH(y_3), \end{aligned}$$

where $\operatorname{sgn} x = 1$, for $x \geq 0$, $\operatorname{sgn} x = -1$ for $x < 0$.

$$(\operatorname{sgn} w_1)(\operatorname{sgn} w_2) = (\operatorname{sgn} x_1)(\operatorname{sgn} x_2),$$

so that the last equation becomes

$$\begin{aligned} (3) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i[t_1 (\log |x_1| - \log |x_3|) + t_2 (\log |x_2| - \log |x_3|)]} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dF(x_1) dF(x_2) dF(x_3) \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i[t_1 (\log |y_1| - \log |y_3|) + t_2 (\log |y_2| - \log |y_3|)]} \operatorname{sgn} y_1 \\ \times \operatorname{sgn} y_2 dH(y_1) dH(y_2) dH(y_3). \end{aligned}$$

Set

$$\begin{aligned} \psi_1(t) &= \int e^{it \log |x|} dF(x); & \varphi_1(t) &= \int e^{it \log |y|} dH(y) \\ \psi_2(t) &= \int e^{it \log |x|} \operatorname{sgn} x dF(x); & \varphi_2(t) &= \int e^{it \log |y|} \operatorname{sgn} y dH(y). \end{aligned}$$

From (3) we have $\psi_2(t_1)\psi_2(t_2)\psi_1(-t_1-t_2) = \varphi_2(t_1)\varphi_2(t_2)\varphi_1(-t_1-t_2)$ for all t_1, t_2 , and from (2) we have $\psi_1(t) = \varphi_1(t)$ for all t , hence, if $\psi_1(-t_1-t_2) \neq 0$, $\psi_2(t_1)\psi_2(t_2) = \varphi_2(t_1)\varphi_2(t_2)$. By hypothesis the zeros of $\psi_1(t)$ are nowhere dense, hence if $\psi_1(-t_1-t_2) = 0$ there is a sequence $t^{(n)}$, such that $t^{(n)} \rightarrow -t_1-t_2$ and $\psi_1(t^{(n)}) \neq 0$. Now take an arbitrary sequence $t_1^{(n)}$ such that $t_1^{(n)} \rightarrow t_1$, then $t_2^{(n)} = -t^{(n)} - t_1^{(n)}$ must tend to t_2 . For each n we have $\psi_2(t_1^{(n)})\psi_2(t_2^{(n)}) = \varphi_2(t_1^{(n)})\varphi_2(t_2^{(n)})$. All the functions appearing are continuous, thus we see that $\psi_2(t_1)\psi_2(t_2) = \varphi_2(t_1)\varphi_2(t_2)$ for all t_1, t_2 . From this it follows directly that either $\psi_2(t) = \varphi_2(t)$ for all t or $\psi_2(t) = -\varphi_2(t)$ for all t . We have³

$$\begin{aligned} \psi_1(t) &= \int_0^{\infty} e^{it \log x} dF(x) + \int_{-\infty}^0 e^{it \log (-x)} dF(x) \\ \psi_2(t) &= \int_0^{\infty} e^{it \log x} dF(x) - \int_{-\infty}^0 e^{it \log (-x)} dF(x) \end{aligned}$$

³ The assumption has been made implicitly that $F(x)$ and $G(x)$ are continuous at $x = 0$, otherwise the distribution of x_i/x_n is not properly defined, and the functions $\varphi_i(t)$ and $\psi_i(t)$ are then not defined. Similar assumptions will be made whenever necessary in later theorems.

$$\varphi_1(t) = \int_0^\infty e^{it \log x} dH(x) + \int_{-\infty}^0 e^{it \log(-x)} dH(x)$$

and
$$\varphi_2(t) = \int_0^\infty e^{it \log x} dH(x) - \int_{-\infty}^0 e^{it \log(-x)} dH(x).$$

Combining these expressions with the relations obtained above leads, by Fourier inversion, to the result that either $F(x) \equiv H(x)$ or $H(x) \equiv 1 - F(-x)$. We have shown that either y or $-y$ has the same distribution as x , that is, either $e^{-a}x'$ or $-e^{-a}x'$ has the same distribution as x .

Theorem Ib states essentially that the joint distribution of the set x_α/x_n , $\alpha = 1, \dots, n-1$, determines the distribution of x except for a scale parameter and possibly a reflection. In the event that x has an asymmetrical distribution, and if it is desired to rule out negative changes of scale, a variation of this procedure is necessary. The next result is appropriate for this situation.

THEOREM Ic: Let x have distribution $F(x)$ such that the zeros of $\int e^{it \log|x|} dF(x)$ are nowhere dense, and let x' have distribution $G(x')$. Let x_1, \dots, x_n and x'_1, \dots, x'_n be samples from the distributions of x and x' , with $n \geq 3$. Express x_1, \dots, x_n and x'_1, \dots, x'_n in spherical coordinates

$$\begin{aligned} x_1 &= r \cos \theta_1, & x'_1 &= r' \cos \theta'_1 \\ x_2 &= r \sin \theta_1 \cos \theta_2, & x'_2 &= r' \sin \theta'_1 \cos \theta'_2 \\ &\vdots & &\vdots \\ x_n &= r \sin \theta_1 \sin \theta_2 \dots \sin \theta_{n-1}, & x'_n &= r' \sin \theta'_1 \sin \theta'_2 \dots \sin \theta'_{n-1}. \end{aligned}$$

Then $\theta_1, \dots, \theta_{n-1}$ have the same joint distribution as $\theta'_1, \dots, \theta'_{n-1}$ if and only if there exists a positive constant k such that kx' and x have the same distribution.

PROOF: Sufficiency of the condition is an immediate consequence of the fact that $\theta_1, \dots, \theta_{n-1}$ are invariant under the transformation $x = kx'$, with $k > 0$. If $\theta_1, \dots, \theta_{n-1}$ have the same joint distribution as $\theta'_1, \dots, \theta'_{n-1}$ then the set $\{x_\alpha/x_n\}$ have the same joint distribution as the set $\{x'_\alpha/x'_n\}$, hence, by Theorem Ib, there exists a constant c such that cx' has the same distribution as x . To establish necessity of the condition we must show that $|c|x'$ has the same distribution as x .

Set $y = |c|x'$, and let y_1, \dots, y_n be expressed in spherical coordinates; y_1, \dots, y_n have the same angular coordinates $\theta'_1, \dots, \theta'_{n-1}$. This implies that x_1/r and x_2/r have the same joint distribution as y_1/R and y_2/R , where $R = \sqrt{y_1^2 + \dots + y_n^2}$; $\frac{x_1}{r} \Big/ \frac{x_2}{r} = x_1/x_2$, therefore x_1/x_2 has the same distribution as y_1/y_2 , so that

$$\int \int e^{it \log \frac{x_1}{x_2}} \operatorname{sgn} \left(\frac{x_1}{x_2} \right) dF(x_1) dF(x_2) = \int \int e^{it \log \frac{y_1}{y_2}} \operatorname{sgn} \left(\frac{y_1}{y_2} \right) dH(y_1) dH(y_2)$$

if y has distribution $H(y)$. $\text{Sgn} \left(\frac{x_1}{|x_2|} \right) = \text{sgn } x_1$, so that the last equation yields

$$\begin{aligned} \int_{-\infty}^{\infty} e^{it \log |x|} \text{sgn } x dF(x) \cdot \int_{-\infty}^{\infty} e^{-it \log |x|} dF(x) \\ = \int_{-\infty}^{\infty} e^{it \log |x|} \text{sgn } x dH(x) \cdot \int_{-\infty}^{\infty} e^{-it \log |x|} dH(x). \end{aligned}$$

We know already that $|x|$ and $|y|$ have the same distribution, so that

$$(4) \quad \int_{-\infty}^{\infty} e^{it \log |x|} dF(x) = \int_{-\infty}^{\infty} e^{it \log |x|} dH(x),$$

thus

$$(5) \quad \int_{-\infty}^{\infty} e^{it \log |x|} \text{sgn } x dF(x) = \int_{-\infty}^{\infty} e^{it \log |x|} \text{sgn } x dH(x),$$

except possibly for zeros of $\int_{-\infty}^{\infty} e^{-it \log |x|} dF(x)$. By hypothesis the exceptional points are nowhere dense, so that, by continuity, (5) holds for all t . (4) and (5) together imply, as in the proof of Theorem Ib, that $F(x) \equiv H(x)$, i.e., x and $|c| x'$ have the same distribution.

The next three results are generalizations of Theorems Ia, b, c, to analogous multivariate situations. The first of these is a direct generalization of Theorem Ia.

THEOREM IIa: Let x_1, \dots, x_k have joint distribution $F(x_1, \dots, x_k)$ such that the complement of the set S of zeros of $\int e^{i \sum t_r x_r} dF(x_1, \dots, x_k)$ is ϵ -connected, where ϵ is the g.l.b. of $|t|$ for (t) in S , and let y_1, \dots, y_k have joint distribution $G(y_1, \dots, y_k)$. Let $(x_1^\alpha, \dots, x_k^\alpha)$ and $(y_1^\alpha, \dots, y_k^\alpha)$, $\alpha = 1, \dots, n$, be samples from these distributions, with $n \geq 3$. Then $w_{i\beta} = x_i^\beta - x_i^\alpha$, $i = 1, \dots, k$, $\beta = 1, \dots, n-1$, have the same joint distribution as the corresponding set $v_{i\beta} = y_i^\beta - y_i^\alpha$ if and only if there exist constants a_1, \dots, a_k such that $y_1 + a_1, \dots, y_k + a_k$ have the same joint distribution as x_1, \dots, x_k .

PROOF: Set

$$\varphi(t_1, \dots, t_k) = \int e^{i \sum_1^k t_r x_r} dF(x_1, \dots, x_k),$$

$$\psi(t_1, \dots, t_k) = \int e^{i \sum_1^k t_r y_r} dG(y_1, \dots, y_k).$$

If $w_{i\beta}$, $i = 1, \dots, k$, $\beta = 1, 2$, have the same joint distribution as $v_{i\beta}$, then, as in the proof of Theorem Ia, we have

$$\begin{aligned} (6) \quad \varphi(t_{11}, \dots, t_{k1}) \varphi(t_{12}, \dots, t_{k2}) \varphi(-t_{11} - t_{12}, \dots, -t_{k1} - t_{k2}) \\ = \psi(t_{11}, \dots, t_{k1}) \psi(t_{12}, \dots, t_{k2}) \psi(-t_{11} - t_{12}, \dots, -t_{k1} - t_{k2}). \end{aligned}$$

Again, as before, $|\varphi| = |\psi|$; $\varphi(t_1, \dots, t_k)$ and $\psi(t_1, \dots, t_k)$ are continuous; $\varphi(0, 0, \dots, 0) = \psi(0, 0, \dots, 0) = 1$. There will exist a neighborhood N of $(0, 0, \dots, 0)$ such that for $(t_1, \dots, t_k) \in N$ the function $f(t_1, \dots, t_k) = \frac{\varphi(t_1, \dots, t_k)}{\psi(t_1, \dots, t_k)}$ is defined and continuous. Then there will exist a neighborhood $N' \subset N$ such that in N' there exists a uniquely determined branch of $\arg f(t_1, \dots, t_k)$, continuous in N' , and such that if $(t_1, \dots, t_k) \in N'$ and $(u_1, \dots, u_k) \in N'$ then $\arg f(t_1 + u_1, \dots, t_k + u_k)$ is also uniquely determined and continuous. For $(t) \in N'$ and $(u) \in N'$, $\arg f$ satisfies the relation

$$\arg f(t_1, \dots, t_k) + \arg f(u_1, \dots, u_k) = \arg f(t_1 + u_1, \dots, t_k + u_k).$$

It is easily shown that any continuous function satisfying the equation above must be of the form $\Sigma a_r t_r$, therefore

$$(7) \quad \varphi(t_1, \dots, t_k) = e^{i \sum_1^k a_r t_r} \psi(t_1, \dots, t_k); \quad (t) \in N'.$$

Just as in the proof of Ia the relation (7) may be extended, by use of (6), to hold for all t . This implies, finally, that the set $\{y_i + a_i\}$ have the same joint distribution as the set $\{x_i\}$.

Theorem IIb is a generalization of Theorem Ib to multivariate distributions.

THEOREM IIb: Let x_1, \dots, x_k have distribution $F(x_1, \dots, x_k)$ such that the zeros of $\int e^{i \Sigma t_r \log |x_r|} dF(x_1, \dots, x_k)$ are nowhere dense, and let y_1, \dots, y_k have distribution $G(y_1, \dots, y_k)$. Let $(x_1^\alpha, \dots, x_k^\alpha)$ and $(y_1^\alpha, \dots, y_k^\alpha)$, $\alpha = 1, \dots, n$, be samples, with $n \geq 3$. Then the set $w_{i\beta} = x_i^\beta / x_i^n$, $i = 1, \dots, k$, $\beta = 1, \dots, n - 1$, have the same joint distribution as the corresponding set $v_{i\beta} = y_i^\beta / y_i^n$ if and only if there exist constants c_1, \dots, c_k such that the set $c_i y_i$ have the same distribution as the x_i .

PROOF: The demonstration is parallel to that of Theorem Ib. By Theorem IIa there exist a_1, \dots, a_k such that

$$E(e^{i \Sigma t_r \log |x_r|}) = E(e^{i \Sigma t_r (\log |y_r| + a_r)}).$$

Set $z_r = e^{a_r} y_r$, then

$$(8) \quad \int e^{i \Sigma t_r \log |x_r|} dF(x_1, \dots, x_k) = \int e^{i \Sigma t_r \log |z_r|} dH(z_1, \dots, z_k),$$

where (z_1, \dots, z_k) have distribution function $H(z_1, \dots, z_k)$.

We shall continue the proof from here under the assumption that $k = 2$. It will be evident how the proof goes for any k . We have, since z_r^β / z_i^3 have the same joint distribution as x_r^β / x_i^3 ,

$$(9) \quad \begin{aligned} & \int \int_{-\infty}^{\infty} e^{i \Sigma t_r \beta (\log |x_r^\beta| - \log |x_i^3|)} \operatorname{sgn} \left(\frac{x_1^1}{x_1^3} \right) \operatorname{sgn} \left(\frac{x_2^1}{x_1^3} \right) dF(x_1^1, x_2^1) dF(x_1^2, x_2^2) dF(x_1^3, x_2^3) \\ &= \int \int_{-\infty}^{\infty} e^{i \Sigma t_r \beta (\log |z_r^\beta| - \log |z_i^3|)} \operatorname{sgn} \left(\frac{z_1^1}{z_1^3} \right) \operatorname{sgn} \left(\frac{z_2^1}{z_1^3} \right) dH(z_1^1, z_2^1) dH(z_1^2, z_2^2) dH(z_1^3, z_2^3). \end{aligned}$$

Both members of (9) are evaluated as products, just as was done in previous proofs, and from the result, combined with (8), we conclude, as in Theorem Ib, that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_1 dF(x_1, x_2) = s_1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_1 dH(x_1, x_2),$$

where $s_1 = \pm 1$, for all (t_1, t_2) . Similarly

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_2 dF(x_1, x_2) = s_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_2 dH(x_1, x_2)$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dF(x_1, x_2) = s_3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dH(x_1, x_2),$$

with $s_2 = \pm 1$, $s_3 = \pm 1$.

$$\text{Set} \quad \varphi_1(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_1 dF(x_1, x_2)$$

$$\varphi_2(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_2 dF(x_1, x_2)$$

$$\varphi_{12}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\sum t_r \log |x_r|} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dF(x_1, x_2)$$

and let $\psi_1(t_1, t_2)$, $\psi_2(t_1, t_2)$, and $\psi_{12}(t_1, t_2)$ denote the corresponding transforms of $H(x_1, x_2)$. We have

$$(10) \quad \begin{cases} \varphi_1(t_1, t_2) = s_1 \psi_1(t_1, t_2) \\ \varphi_2(t_1, t_2) = s_2 \psi_2(t_1, t_2) \\ \varphi_{12}(t_1, t_2) = s_3 \psi_{12}(t_1, t_2) \end{cases}$$

with $s_1 = \pm 1$, $s_2 = \pm 1$, and $s_3 = \pm 1$.

Now, as in (9), by considering $E \left[e^{i\sum t_r \beta (\log |x_r^2| - \log |x_r^3|)} \operatorname{sgn} \left(\frac{x_1^1}{x_1^3} \right) \operatorname{sgn} \left(\frac{x_2^2}{x_2^3} \right) \right]$ we obtain the relation

$$\begin{aligned} \varphi_1(t_{11}, t_{21}) \varphi_2(t_{12}, t_{22}) \varphi_{12}(-t_{11} - t_{12}, -t_{21} - t_{22}) \\ = \psi_1(t_{11}, t_{21}) \psi_2(t_{12}, t_{22}) \psi_{12}(-t_{11} - t_{12}, -t_{21} - t_{22}), \end{aligned}$$

showing that s_1, s_2, s_3 , may be chosen so that $s_1 s_2 s_3 = 1$, that is, $s_1 s_2 = s_3$.

Consider now the variates $z'_r = s_r z_r$, $r = 1, 2$. Let $K(z'_1, z'_2)$ be the distribution function of z'_1, z'_2 . If we let $\theta_1(t_1, t_2)$, $\theta_2(t_1, t_2)$, and $\theta_{12}(t_1, t_2)$ be the transforms of K which correspond to $\varphi_1(t_1, t_2)$, $\varphi_2(t_1, t_2)$, and $\varphi_{12}(t_1, t_2)$ respectively, it is evident that

$$(11) \quad \begin{aligned} \varphi_1(t_1, t_2) &= \theta_1(t_1, t_2) \\ \varphi_2(t_1, t_2) &= \theta_2(t_1, t_2) \\ \varphi_{12}(t_1, t_2) &= \theta_{12}(t_1, t_2). \end{aligned}$$

Moreover, from (8),

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i \sum t_r \log |z_r|} dF(x_1, x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i \sum t_r \log |z_r|} dK(x_1, x_2).$$

The last relation, together with the equations (11) imply that $F(x)$ and $K(x)$ coincide in each quadrant, thus $F(x_1, x_2) \equiv K(x_1, x_2)$ for all x_1, x_2 .

The final result is that z'_1, z'_2 have the same distribution as x_1, x_2 , i.e., $s_1 e^{a_1} y_1$ and $s_2 e^{a_2} y_2$ have the same joint distribution as x_1 and x_2 .

The next result bears the same relation to Theorem IIb that Theorem Ic bears to Theorem Ib, that is, only *positive* scale changes are to be permitted.

THEOREM IIC: Let x_1, \dots, x_k have distribution $F(x_1, \dots, x_k)$ such that the zeros of $\int e^{i \sum t_r \log |z_r|} dF(x_1, \dots, x_k)$ are nowhere dense, and let y_1, \dots, y_k have distribution $G(y_1, \dots, y_k)$. Let $(x_1^\alpha, \dots, x_k^\alpha)$ and $(y_1^\alpha, \dots, y_k^\alpha)$, $\alpha = 1, 2, \dots, n$, be samples with $n \geq 3$. Express $x_1^\alpha, \dots, x_k^\alpha$ and $y_1^\alpha, \dots, y_k^\alpha$ in spherical coordinates

$$\begin{aligned} x_i^1 &= r_i \cos \theta_i^1, & y_i^1 &= R_i \cos \varphi_i^1, \\ x_i^2 &= r_i \sin \theta_i^1 \cos \theta_i^2, & y_i^2 &= R_i \sin \varphi_i^1 \cos \varphi_i^2, \\ &\vdots & &\vdots \\ x_i^n &= r_i \sin \theta_i^1 \dots \sin \theta_i^{n-1}, & y_i^n &= R_i \sin \varphi_i^1 \dots \sin \varphi_i^{n-1}. \end{aligned}$$

Then $\{\theta_i^\beta\}$, $i = 1, \dots, k$, $\beta = 1, \dots, n-1$, have the same joint distribution as $\{\varphi_i^\beta\}$ if and only if there exist constants $k_i > 0$, $i = 1, \dots, k$, such that the set $k_i y_i$ have the same joint distribution as the set x_i .

PROOF: If $\{\theta_i^\beta\}$ have the same distribution as $\{\varphi_i^\beta\}$ then it follows that $\left\{ \frac{x_i^\beta}{x_i^n} \right\}$ have the same distribution as $\left\{ \frac{y_i^\beta}{y_i^n} \right\}$, hence by Theorem IIb there exist constants c_i such that $\{c_i y_i\}$ have the same distribution as $\{x_i\}$. Set $z_i = |c_i| y_i$; we wish to show that $\{z_i\}$ have the same distribution as $\{x_i\}$. By equation (8) in Theorem IIb it is known that $\{ |z_i| \}$ have the same distribution as $\{ |x_i| \}$, moreover, if we express z_i^α in spherical coordinates, the angular coordinates are

the same as those of y_i^α , therefore $\left\{\frac{x_i^1}{x_i^2}\right\}$ have the same distribution as $\left\{\frac{z_i^1}{z_i^2}\right\}$, since these functions are obtainable in terms of the angular coordinates.

As before, we shall continue the proof from here under the assumption that $k = 2$. The procedure is a generalization of the procedure in the proof of Theorem Ic. $\operatorname{sgn} x_i^1 = \operatorname{sgn} \left\{\frac{x_i^1}{x_i^2}\right\}$, and similarly for y , therefore

$$(12) \quad \begin{aligned} & \int \int e^{i \sum_{r=1}^2 t_r (\log |z_r^1| - \log |z_r^2|)} \operatorname{sgn} x_i^1 dF(x_1^1, x_2^1) dF(x_1^2, x_2^2) \\ &= \int \int e^{i \sum_{r=1}^2 t_r (\log |z_r^1| - \log |z_r^2|)} \operatorname{sgn} x_i^1 dH(x_1^1, x_2^1) dH(x_1^2, x_2^2), \quad i = 1, 2, \end{aligned}$$

where it is assumed that z_1, z_2 have distribution $H(z_1, z_2)$. As before, set

$$\begin{aligned} \varphi(t_1, t_2) &= \int e^{i \sum_{r=1}^2 t_r \log |z_r|} dF(x_1, x_2), \\ \varphi_i(t_1, t_2) &= \int e^{i \sum_{r=1}^2 t_r \log |z_r|} \operatorname{sgn} x_i dF(x_1, x_2), \quad i = 1, 2, \\ \varphi_{12}(t_1, t_2) &= \int e^{i \sum_{r=1}^2 t_r \log |z_r|} \operatorname{sgn} x_1 \operatorname{sgn} x_2 dF(x_1, x_2), \end{aligned}$$

and denote the corresponding transforms of $H(x_1, x_2)$ by $\theta(t_1, t_2)$, $\theta_1(t_1, t_2)$, $\theta_2(t_1, t_2)$, and $\theta_{12}(t_1, t_2)$. It has been remarked already that $\{|z_i|\}$ have the same distribution as $\{|x_i|\}$, therefore $\theta(t_1, t_2) = \varphi(t_1, t_2)$. Equation (12) yields the relation $\varphi_i(t_1, t_2)\varphi(-t_1, -t_2) = \theta_i(t_1, t_2)\theta(-t_1, -t_2)$, $i = 1, 2$; the zeros of $\varphi(t_1, t_2)$ are nowhere dense, so that it can be concluded that $\varphi_i(t_1, t_2) = \theta_i(t_1, t_2)$, $i = 1, 2$. Now, from an equation similar to (12) we obtain $\varphi_{12}(t_1, t_2) = \theta_{12}(t_1, t_2)$. As in Theorem IIb, the four relations above together imply that $F(x_1, x_2) \equiv H(x_1, x_2)$, in other words, $\{|c_i|y_i\}$ have the same distribution as $\{x_i\}$.

We are now in a position to combine some of the preceding theorems so as to obtain analogous results for scale and location parameters together.

THEOREM IIIA: Let x have distribution $F(x)$ such that the zeros of $\int e^{itz} dF(x)$ satisfy the condition of Theorem Ia, and the zeros of

$$\int \int \int e^{it_1 \log |x_1 - x_2| + it_2 \log |x_2 - x_3|} dF(x_1) dF(x_2) dF(x_3)$$

are nowhere dense, and let y have distribution $G(y)$. Let x_1, \dots, x_n and y_1, \dots, y_n be samples, with $n \geq 9$. Then $w_\alpha = \frac{x_\alpha - x_n}{x_{n-1} - x_n}$, $\alpha = 1, \dots, n-2$,

have the same joint distribution as the corresponding set $w'_\alpha = \frac{y_\alpha - y_n}{y_{n-1} - y_n}$ if and only if there exist constants a, c , such that $c(y - a)$ and x have the same distribution.

PROOF: Sufficiency of the condition is an immediate consequence of the fact that w'_α is invariant under transformations of the form $y' = c(y - a)$. Assume then that $\{w_\alpha\}$ and $\{w'_\alpha\}$ have the same joint distribution. By elementary transformations it is evident that the functions $\frac{x_1 - x_3}{x_7 - x_9}, \frac{x_4 - x_6}{x_7 - x_9}, \frac{x_2 - x_3}{x_8 - x_9}, \frac{x_5 - x_6}{x_8 - x_9}$ have the same joint distribution as the corresponding functions of the y 's, if $n \geq 9$. Since x_1, \dots, x_n form a sample it follows that the pairs $\{x_1 - x_3, x_2 - x_3\}$, $\{x_4 - x_6, x_5 - x_6\}$, $\{x_7 - x_9, x_8 - x_9\}$, have the same joint distributions and are pairwise independent, and similarly for the corresponding functions of the y 's. Theorem IIb assures the existence of constants c_1, c_2 , such that $c_1(y_1 - y_3), c_2(y_2 - y_3)$ have the same joint distribution as $(x_1 - x_3), (x_2 - x_3)$. Considering separately the marginal distributions it is seen that $c_1(y_1 - y_3)$ has the same distribution as $c_2(y_2 - y_3)$. $y_1 - y_3$ and $y_2 - y_3$ have the same distribution, therefore either $c_2 = c_1$, or $c_2 = -c_1$. Set $u_\alpha = x_\alpha - x_3$, $v_\alpha = c_1(y_\alpha - y_3)$, $\alpha = 1, 2$. We have, for the distributions of (u_1, u_2) and (v_1, v_2) , relations corresponding to (10) in Theorem IIb, with the additional condition that $s_1 = s_2$, because of the symmetry in the variables. This implies that either (v_1, v_2) or $(-v_1, -v_2)$ have the same joint distribution as (u_1, u_2) , that is, there exists c such that $c(y_1 - y_3)$ and $c(y_2 - y_3)$ have the same joint distribution as $x_1 - x_3$ and $x_2 - x_3$. Application of Theorem Ia now completes the proof.

Just as before, there is an analogous situation when we consider angular coordinates instead of quotients. The proof is immediate; the angular coordinates determine the angular coordinates of $\{x_1 - x_3, x_2 - x_3\}$, $\{x_4 - x_6, x_5 - x_6\}$, and $\{x_7 - x_9, x_8 - x_9\}$, arranged as a sample. Then the constants c_1, c_2 in the proof of Theorem IIIa are both positive; it follows that $c_1 = c_2$. Application of Theorem Ia gives

THEOREM IIIb: Let x_1, \dots, x_n and y_1, \dots, y_n satisfy the hypotheses of Theorem IIIa. Set

$$\begin{aligned} x_1 - x_n &= r \cos \theta_1, & y_1 - y_n &= r' \cos \theta'_1, \\ x_2 - x_n &= r \sin \theta_1 \cos \theta_2, & y_2 - y_n &= r' \sin \theta'_1 \cos \theta'_2, \end{aligned}$$

$$x_{n-1} - x_n = r \sin \theta_1 \cdots \sin \theta_{n-2}; \quad y_{n-1} - y_n = r' \sin \theta'_1 \cdots \sin \theta'_{n-2}.$$

Then $\theta_1, \dots, \theta_{n-2}$ have the same joint distribution as $\theta'_1, \dots, \theta'_{n-2}$ if and only if there exist constants a and $c > 0$ such that $c(y - a)$ has the same distribution as x .

Theorem IVa is a generalization of Theorem Ia to cover arbitrary linear combinations of some subset of the sample.

THEOREM IVa: Suppose x has distribution $F(x)$ such that $\int e^{itz} dF(x)$ does not vanish, and let y have distribution $G(y)$. Consider the functions $w_\alpha = x_\alpha - \sum_{\beta=1}^{n-m} l_{\alpha\beta} x_{m+\beta}$, $w'_\alpha = y_\alpha - \sum_{\beta=1}^{n-m} l_{\alpha\beta} y_{m+\beta}$, $\alpha = 1, 2, \dots, m$, $\beta = 1, 2, \dots$;

$n - m$, and suppose that $m > n - m$. Then, if $\{w_\alpha\}$ have the same joint distribution as $\{w'_\alpha\}$ and if $\sum_{\beta=1}^{n-m} l_{\alpha\beta} \neq 1$ for some α , it follows that $F(y) \equiv G(y)$; if $\sum_{\beta} l_{\alpha\beta} = 1$ for all α there exists a constant a such that $F(y - a) \equiv G(y)$.

PROOF: Denote the characteristic functions of x and y by $\varphi(t)$ and $\psi(t)$ respectively. By expressing the fact that $\{w_\alpha\}$ and $\{w'_\alpha\}$, $\alpha = 1, 2, \dots, n - m + 1$, have the same characteristic function we obtain the functional equation

$$\prod_{\alpha=1}^{n-m+1} \varphi(t_\alpha) \prod_{\beta=1}^{n-m} \varphi\left(-\sum_{\alpha=1}^{n-m+1} l_{\alpha\beta} t_\alpha\right) = \prod_{\alpha=1}^{n-m+1} \psi(t_\alpha) \prod_{\beta=1}^{n-m} \psi\left(-\sum_{\alpha=1}^{n-m+1} l_{\alpha\beta} t_\alpha\right).$$

By hypothesis $\varphi(t)$ does not vanish, therefore $\psi(t)$ has no zeros, because of the relation above. $\varphi(t)$ and $\psi(t)$ are continuous, thus the function $f(t) = \log \varphi(t) - \log \psi(t)$ can be uniquely defined in a continuous manner for all t . The equation above becomes

$$(13) \quad \sum_{\alpha=1}^{n-m+1} f(t_\alpha) + \sum_{\beta=1}^{n-m} f\left(-\sum_{\alpha=1}^{n-m+1} l_{\alpha\beta} t_\alpha\right) = 0.$$

The constants $l_{\alpha\beta}$ are necessarily linearly dependent, so that, for some α , $l_{\alpha\beta}$ can be expressed as a linear combination of the others; suppose then that

$$l_{n-m+1,\beta} = \sum_{\alpha=1}^{n-m} e_\alpha l_{\alpha\beta}.$$

Putting these values in (13) we have

$$(14) \quad \sum_{\alpha=1}^{n-m+1} f(t_\alpha) + \sum_{\beta=1}^{n-m} f\left(-\sum_{\alpha=1}^{n-m} l_{\alpha\beta}(t_\alpha + t_{n-m+1}e_\alpha)\right) = 0.$$

It can be assumed that $\sum e_\alpha^2 \neq 0$, for, if $e_\alpha = 0$ for all α , we have $l_{n-m+1,\beta} = 0$, $\beta = 1, \dots, n - m$, that is, $w'_{n-m+1} = y_{n-m+1}$ and $w_{n-m+1} = x_{n-m+1}$, hence x and y have the same distribution. Assuming $e_1 \neq 0$, set $t_\alpha = -e_\alpha t_{n-m+1}$, $\alpha = 2, \dots, n - m$, in (14), obtaining

$$(15) \quad f(t_1) + \sum_{\alpha=2}^{n-m} f(-e_\alpha t_{n-m+1}) + f(t_{n-m+1}) + \sum_{\beta=1}^{n-m} f(-l_{1\beta}(t_1 + e_1 t_{n-m+1})) = 0,$$

now, recalling that $f(0) = 0$, set $t_{n-m+1} = 0$, getting $f(t_1) + \sum_{\beta=1}^{n-m} f(-l_{1\beta}t_1)$.

Evaluating this with argument $t_1 + e_1 t_{n-m+1}$, and substituting back in (15) it appears that

$$(16) \quad f(t_1) + f(t_{n-m+1}) + \sum_{\alpha=2}^{n-m} f(-e_\alpha t_{n-m+1}) = f(t_1 + e_1 t_{n-m+1}).$$

Now setting $t_1 = 0$ in (16) we have the relation

$$f(t_{n-m+1}) + \sum_{\alpha=2}^{n-m} f(-e_\alpha t_{n-m+1}) = f(e_1 t_{n-m+1}).$$

thus we have finally $f(t_1) + f(e_1 t_{n-m+1}) = f(t_1 + e_1 t_{n-m+1})$, or, since $e_1 \neq 0$, $f(t_1 + t_2) = f(t_1) + f(t_2)$. The last relation implies that $f(t) = ct$, since $f(t)$ is continuous. Now replace $f(t)$ by ct in (13), getting $c \left\{ \sum_{\alpha=1}^{n-m+1} t_\alpha - \sum_{\alpha=1}^{n-m+1} \sum_{\beta=1}^{n-m} l_{\alpha\beta} t_\alpha \right\} = 0$, that is, either $c = 0$, or $\sum_{\beta=1}^{n-m} l_{\alpha\beta} = 1$ for all α . We conclude then that $\varphi(t) = \psi(t)$, unless $\sum_{\beta} l_{\alpha\beta} = 1$ for all α . If $\sum_{\beta} l_{\alpha\beta} = 1$ for all α we have $\varphi(t) = e^{ct} \psi(t)$. $\varphi(-t) = \overline{\varphi(t)}$ and $\psi(-t) = \overline{\psi(t)}$, hence c is of the form $c = ia$, where a is real, in other words $\varphi(t) = e^{iat} \psi(t)$, thus concluding the proof of the theorem.

It was assumed in Theorem IVa that $\varphi(t)$ has no zeros. If $\varphi(t)$ has zeros we have proved that, for an interval $|t| < \epsilon$, $\varphi(t) = \psi(t)$ (or $\varphi(t) = e^{iat} \psi(t)$). This does not necessarily imply the result of Theorem IVa, but it does imply at least that if the k th moments of x and of y (or of $y - a$) both exist they are equal.

The last result in this series can be proved by methods similar to those used in Theorem IVa.

THEOREM IVB: Let x and y satisfy the hypotheses of Theorem IVa. Suppose, moreover, that $m > 2(n - m)$, that the rank of $\|l_{\alpha\beta}\|$ is $n - m$, and that $\sum_{\beta=1}^{n-m} l_{\alpha\beta} \neq 1$ for at least $2m - n$ values of α . Then, if there exist constants $\{c_\alpha\}$ such that the set $\{c_\alpha w'_\alpha\}$ have the same joint distribution as $\{w_\alpha\}$, it follows that, for some α , $c_\alpha y$ has the same distribution as x .

3. Application to Composite Hypotheses. The results of section 2 have a significant application in the theory of testing composite hypotheses. Suppose that x has a distribution of the form $F(x, \theta_1, \theta_2)$, and that the hypothesis $\theta_2 = \theta_2^0$ is to be tested, without reference to the value of θ_1 . We assume that the parameters are independent, i.e., $F(x, \theta_1, \theta_2) \equiv F(x, \theta_1', \theta_2')$ implies that $\theta_1 = \theta_1'$ and $\theta_2 = \theta_2'$. It is true in a wide class of important cases that, given a sample x_1, \dots, x_n from the distribution $F(x, \theta_1, \theta_2)$, there exist functions $y_\alpha(x_1, \dots, x_n)$, $\alpha = 1, 2, \dots, p$, such that $\{y_\alpha\}$ have joint distribution independent of θ_1 , but depending on θ_2 . Now if the $\{y_\alpha\}$ are such that their joint distribution redetermines the original distribution, except for θ_1 , one can reasonably use the p -dimensional distribution of the $\{y_\alpha\}$ for testing the hypothesis $\theta_2 = \theta_2^0$, thus reducing the composite hypothesis to a simple hypothesis. In testing this simple hypothesis, every alternative hypothesis (corresponding to a value of θ_2) determines a distribution of x among the alternatives $F(x, \theta_1, \theta_2)$ except for the unknown θ_1 , that is, there is a one-to-one correspondence between the two sets of alternative hypotheses, expressed by the fact that if $\theta_2' = \theta_2''$ then the distributions of the set $\{y_\alpha\}$ corresponding to $\theta_2 = \theta_2'$ and $\theta_2 = \theta_2''$ must be different.

Suppose, for example, that it is desired to test whether $y = x - a$ for some a has the distribution $F(y, \theta^0)$, with the assumption that, for some a , y has the

distribution $F(y, \theta)$. Given a sample one can form the set $w_\alpha = x_\alpha - x_n$, $\alpha = 1, 2, \dots, n-1$, obtaining the distribution $G(w_1, \dots, w_{n-1}, \theta)$; now consider the simple hypothesis $\theta = \theta^0$, knowing that G determines θ , by Theorem Ia. Similarly one can test whether cx , for some $c \neq 0$, has distribution $F(y, \theta^0)$, by forming $w_\alpha = x_\alpha/x_n$, $\alpha = 1, \dots, n-1$, or by expressing (x_1, \dots, x_n) in spherical coordinates and considering the angular coordinates, according to whether both positive and negative or only positive values of c are to be allowed.

In the same way one can test the hypothesis $\theta = \theta^0$ under the assumption that $c(x-a)$ has distribution $F(y, \theta)$ by forming $w_\alpha = \frac{x_\alpha - x_n}{x_{n-1} - x_n}$, $\alpha = 1, \dots, n-2$, or by expressing $(x_1 - x_n, \dots, x_{n-1} - x_n)$ in spherical coordinates and considering the angular coordinates.

Theorem IVa may be applied to analogous problems, in which the hypothesis $\theta = \theta^0$ is to be tested under the assumption that $y = u - \sum a_i x_i$ has distribution $F(y, \theta)$ for fixed values of the x_i , with the a_i unknown. In such problems there exist linear combinations of the observed values of y which are independent of the a_i . By Theorem IVa, under certain conditions the joint distribution of these linear combinations determines the original distribution of y , without regard to the a_i .

In applying some of the preceding results we must verify in certain cases that the zeros of $\int e^{itz} dF(x)$ are nowhere dense, for a certain distribution function.

By a change of variable the condition of Theorem Ib can be stated in this form; moreover if $F(x)$ satisfies this condition it is evident that it satisfies the condition of Theorem Ia. A sufficient condition applicable to a considerable class of cases has been obtained by Levinson [4]; if $f(x)$ is $O(e^{-\theta(x)})$ as $x \rightarrow \infty$, where $\theta(x)$ is monotone and $\int_1^\infty \frac{\theta(x)}{x^2} dx$ diverges to ∞ , then $\int e^{itz} f(x) dx$ cannot vanish on an interval without vanishing identically. It is evident that it is likewise sufficient if the corresponding condition holds as $x \rightarrow -\infty$ instead of $+\infty$. In particular, if there exists A such that $f(x) = 0$ for $x > A$ (or for $x < A$) it is a consequence of the Levinson result that $\int e^{itz} f(x) dx$ has no intervals of zeros. It can be established easily that if $f(x)$ is majorized by $|x|^{-(1-\epsilon)}$, $\epsilon > 0$, in the neighborhood of the origin, then $\int e^{it \log |x|} f(x) dx$ has no intervals of zeros.

As a simple example consider the rectangular distribution on $(0, 1)$. Let $(x-a)/r$ have this distribution with a unknown, $r > 0$, and suppose that we are interested only in r . Given a sample x_1, \dots, x_n form the functions $y_\alpha = (x_\alpha - x_n)/r$, $\alpha = 1, \dots, n-1$. Set $y_M = \max(y_\alpha, 0)$, $y_L = \min(y_\alpha, 0)$. Then it can be shown that y_1, \dots, y_{n-1} have probability density $(1 - y_M + y_L)$ in the region $-1 \leq y_\alpha \leq 1$, $y_M - y_L \leq 1$, zero elsewhere. $\psi = y_M - y_L$ is of course the quotient of the sample range by r . It can be shown that ψ has

density $n(n-1)(1-\psi)\psi^{n-2}d\psi$. Theorem Ia makes it possible to base any tests not involving a on the distribution of the y_α , since if the y_α have the stated distribution then $(x-a)/r$ for some a must have the rectangular distribution.

Similarly, suppose $y = (x-a)/r$ has the distribution e^{-y} , $y > 0$, for some a, r . Then $w_\alpha = \frac{x_\alpha - x_n}{r}$, $\alpha = 1, 2, \dots, n-1$, have distribution density $\frac{1}{n} e^{-\sum w_\alpha + n w_L}$, where $w_L = \min(0, w_\alpha)$. Again, the latter distribution may be used to estimate r .

Let us examine the distributions of functions of the type considered, in the case of normality. Assume that x_1, \dots, x_n are a sample of n observations from a normal distribution with unit variance and unknown mean. The variables $y_\alpha = x_\alpha - x_1$, $\alpha = 2, \dots, n$, have a joint normal distribution with zero means and matrix of variances and covariances $\|A^{ij}\| = \|1 + \delta_{ij}\|$. Then Theorem Ia shows that if $\{y_\alpha\}$ have this joint distribution then x is normally distributed with unit variance. Note that $\chi_{n-1}^2 \equiv \sum A_{ij} y_i y_j \equiv \sum (x_\alpha - \bar{x})^2$. If we had $x = x'/\sigma$, then $\sum (x'_\alpha - \bar{x}')^2 = \sigma^2 \chi_{n-1}^2$, giving the estimate $\frac{1}{n-1} \sum (x'_\alpha - \bar{x}')^2$ for σ^2 .

There are, of course, many ways in which the matrix $\|A_{ij}\|$ may be transformed into a diagonal matrix in order to obtain a new set of independently distributed variates; one convenient set is the set $\sqrt{\frac{1}{2}} y_2, \sqrt{\frac{1}{2}} (y_3 - \frac{1}{2} y_2), \dots, \sqrt{\frac{n-1}{n}} \left(y_n - \frac{1}{n-1} \sum_{\alpha=2}^{n-1} y_\alpha \right)$. In terms of the original x 's we have $\sqrt{\frac{1}{2}} (x_2 - x_1), \sqrt{\frac{1}{2}} (x_3 - \frac{1}{2}(x_1 + x_2)), \sqrt{\frac{n-1}{n}} \left(x_n - \frac{1}{n-1} \sum_{\alpha=1}^{n-1} x_\alpha \right)$; these functions of the data are independently distributed according to the normal distribution with zero mean and unit variance.

Similarly, in the case of a sample x_1, \dots, x_n from a normal distribution with zero mean and unknown variance, there exists a set of $n-1$ functions with distributions independent of the variance. A convenient set of functions is the set

$$t_m = \frac{\sqrt{m} x_{m+1}}{\sqrt{\sum_{i=1}^m x_i^2}}; \quad m = 1, \dots, n-1.$$

It is known (see Bartlett [1]) that the variables t_m are independently distributed according to student t -distributions with m degrees of freedom respectively. The set t_m determines the set of angular coordinates obtained by expressing x_1, \dots, x_n in spherical coordinates, hence we can conclude, conversely, that if $\{t_m\}$ have this joint distribution then x is normal with mean zero.

Finally we can eliminate both mean and variance. Suppose x_1, \dots, x_n are a sample from some normal distribution. The variables

$$u_m = \sqrt{\frac{m}{m+1}} \left\{ x_{m+1} - \frac{1}{m} \sum_{i=1}^m x_i \right\}, \quad m = 1, 2, \dots, n-1,$$

are normal and independent with mean zero and some variance. Then we have the set

$$t'_r = \frac{\sqrt{r \left(\frac{r+1}{r+2} \right)} \left\{ x_{r+2} - \frac{1}{r+1} \sum_{i=1}^{r+1} x_i \right\}}{\sqrt{\sum_{j=1}^r \frac{j}{j+1} \left\{ x_{j+1} - \frac{1}{j} \sum_{i=1}^j x_i \right\}^2}}, \quad r = 1, \dots, n-2,$$

independently distributed according to t -distributions with r degrees of freedom respectively. It may be convenient for computational purposes to make use of the identity

$$\sum_{j=1}^r \frac{j}{j+1} \left\{ x_{j+1} - \frac{1}{j} \sum_{i=1}^j x_i \right\}^2 \equiv \sum_{j=1}^{r+1} \left(x_j - \frac{1}{r+1} \sum_{i=1}^{r+1} x_i \right)^2 \equiv \sum_{j=1}^{r+1} (x_j - \bar{x}_{(r+1)})^2.$$

We then have

$$t'_r = \frac{\sqrt{r \left(\frac{r+1}{r+2} \right)} (x_{r+2} - \bar{x}_{(r+1)})}{\sqrt{\sum_{i=1}^{r+1} (x_i - \bar{x}_{(r+1)})^2}}, \quad r = 1, \dots, n-2.$$

Now, by Theorem IIIc, we know that if the set $\{t'_r\}$ has this specified distribution then x must be distributed according to some normal distribution. The set $\{t'_r\}$ may be used to test the goodness of fit of the observations to normality, by first adjusting the set $\{t'_r\}$ to a standard basis of comparison, i.e., by considering $F_r(t'_r)$, where F_r is the corresponding cumulative distribution function and then applying, for example, a χ^2 goodness of fit test to these $n-2$ quantities, with respect to the rectangular distribution on $(0, 1)$.

REFERENCES

- [1] M. S. BARTLETT, *Proc. Camb. Phil. Soc.* Vol. 30 (1934), pp. 164-169.
- [2] J. NEYMAN AND E. S. PEARSON, *Biometrika*, Vol. XXA, (1928) pp. 175-240, 264-294.
- [3] R. A. FISHER, *Proc. Roy. Soc. A*, Vol. 144 (1934), pp. 285-307.
- [4] N. LEVINSON, *Proc. London Math. Soc.* Vol. 41 (1936), pp. 393-407.
- [5] E. J. G. PITMAN, *Biometrika* Vol. 30 (1939), pp. 391-421.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

THE SELECTION OF VARIATES FOR USE IN PREDICTION WITH SOME COMMENTS ON THE GENERAL PROBLEM OF NUISANCE PARAMETERS

BY HAROLD HOTELLING

1. Maximum Correlation as a Test. For predicting or estimating a particular variate y there is frequently available an embarrassingly large number of other variates having some correlation with y . For example, in fitting demand functions by means of economic time series, the number of series of observations having some relation to the demand which is sought to be estimated is apt to be very large, whereas the number of good independent observations on each is quite small. The proper coefficients in the regression equation must ordinarily be determined from the observations, and must not exceed in number the observations on each variate. Furthermore, in order to have a measure of error that will make it possible to distinguish real effects from those due to chance, it is necessary that the number of predictors¹ shall be enough less than the number of observations on each variate so that the residual chance variance can be determined with an appropriate degree of accuracy. It is desirable to select a set of predictors yielding estimates of maximum but determinable accuracy, and at the same time to avoid the fallacies of selection among numerous results of that one which appears most significant and treating it as if it were the only one examined.

Considerations other than maximum and determinate accuracy are of practical importance. The labor of calculation by the method of least squares becomes a serious obstacle to the use of the theoretically optimum set of variates when these are very numerous, though the rapid current development of mechanical and electrical devices suitable for these computations offers a hope that the limits now set in practice in this way will soon be considerably increased. Furthermore, predictions or estimates must, as in speculative business or in military activity, be made from moment to moment, often in a rough manner by persons incapable of or averse to using complex formulae, and in such activities frequent revisions of the regression equations must be made to accord with altered conditions. Also, in temporal predictions, the time of availability of

¹ I use this term for what are often called the independent variates in a regression equation, since these ordinarily are not really independent in the probability sense. Similarly I shall call the "dependent" variate the *predictand*. By *prediction* I mean merely the use of regression equations to estimate some unknown variate by means of the values of related variates, without any necessary connotation of temporal order, though the most interesting applications seem for the most part to be those in which we pass from a knowledge of the past to an estimate of the future.

the values of the predictors is important, since an early prediction (e.g. of the size of a harvest) is more valuable than a later one of the same accuracy.

If we make the usual assumption² that the probability distribution of y is, for every set of values of the predictors, normal with a fixed variance σ^2 and an expectation that is a linear function of the predictors, we shall wish to minimize σ^2 subject to appropriate limitations, and this amounts to the same thing as maximizing the multiple correlation ρ of y with the predictors, since $1 - \rho^2$ is the ratio of σ^2 to the total variance of y , which is the same for all sets of predictors. The estimates s and R of σ and ρ obtained from the available sample are of course a different matter. But it is clear that the value of R provides a suitable criterion of choice under the following conditions: We are called upon to choose one among two or more sets, each consisting of a fixed number of predictors; for each predictor we have a known value corresponding to each of the values y_1, \dots, y_N observed for the predictand; and there is no basis for preferring one of these sets to another either in theory, in observations extraneous to those just specified, or in cost or time of availability. In particular, if just one predictor is to be used, that having the highest sample correlation with the predictand should under these conditions be the one adopted. But in making such a choice a test of its accuracy is required, to take account of the possibility that the wrong choice has been made because of chance fluctuations in the sample correlation coefficients.

There are innumerable economic variates available for prediction of business conditions, and most of these are highly correlated with each other. The selection of one business index instead of another for a particular purpose will involve the question which has exhibited the higher correlation with the quantity to be predicted, and consequently the question of the definiteness with which the difference between the calculated correlations can be regarded as significant.

Our problem evidently has a bearing on governmental policy in selecting among the numerous series of data those whose continuation will be most valuable. The high cost of assembling these statistics dictates a careful selection of a limited number of series having little correlation with each others' current values, but with correlations as great as possible with those things whose prediction or estimation is most important.

2. The Choice of one Predictor with Two Available. Let us take first the simplest case, which may be illustrated by a Michigan State College problem of

² We shall not here go into the question of the applicability of these standard assumptions to time series otherwise than to note that some transformations of observations ordered in time are usually necessary and sufficient to obtain quantities satisfying the assumptions so closely that deviations from them cannot be detected. Such transformations include replacing a variate by its logarithm, and eliminating trend and seasonal variations by least squares. In view of the satisfactory adjusted observations found empirically by these and similar methods, the usual objections to studying time series by exact methods seem much exaggerated.

which Dr. W. D. Baten has told me. The ultimate weight of a mature ox is estimated by means of his length at an early age. The question has been raised, however, whether a more accurate prediction might not be made by means of the calf's girth at his heart. Records were at hand of 13 oxen showing their lengths and girths as calves and also their weights when mature. A regression equation involving both length and girth would presumably give greater accuracy than either variate alone; but it appears that those who make the estimates desire a simple formula involving only one variate. Suppose, then, that in such a sample the correlation of weight with length is $r_1 = .7$, that the correlation of weight with girth is $r_2 = .5$, and that the correlation of girth with length is $r_0 = .4$. Is the difference $r_1 - r_2 = .2$ sufficiently great in relation to its sampling errors to warrant the inference that girth is really a better predictor than length, or must the question be left in abeyance until more observations can be accumulated?

A straightforward procedure which would have been used with little question before the advent of modern exact methods is to calculate the asymptotic approximation to the standard error of $r_1 - r_2$ by the differential method, assuming the three variates to have the trivariate normal distribution, and to regard the difference of the correlations as significant if it exceeds a multiple of this standard error determined by the tables of the normal distribution. The calculation of the asymptotic approximation $\sigma_{r_1-r_2}$ may be carried out in the following manner. Let ρ_1 , ρ_2 , and ρ_0 be the population values of r_1 , r_2 , and r_0 respectively. Then if σ_{ij} denote the population covariance of x_i and x_j ($i, j = 0, 1, 2$), we have

$$\rho_1 = \frac{\sigma_{01}}{\sqrt{\sigma_{00}\sigma_{11}}},$$

with similar formulae for ρ_2 and ρ_0 . Likewise the sample estimates of these parameters are given by such expressions as

$$r_1 = \frac{s_{01}}{\sqrt{s_{00}s_{11}}}.$$

Taking the logarithm of this last expression, expanding about the population values, denoting by the operator δ the deviation of sample from population values of the covariances, and the resultant deviation in r_1 , and dropping terms of order higher than the first, we have:

$$\delta r_1 = \rho_1 \left(\frac{\delta s_{01}}{s_{01}} - \frac{\delta s_{00}}{2s_{00}} - \frac{\delta s_{11}}{2s_{11}} \right).$$

In the same way

$$\delta r_2 = \rho_2 \left(\frac{\delta s_{02}}{s_{02}} - \frac{\delta s_{00}}{2s_{00}} - \frac{\delta s_{22}}{2s_{22}} \right).$$

The asymptotic value of the sampling covariance is obtained by multiplying these two expressions together and taking the expectation. The sampling covariance of two estimates of covariance of the usual kind (sum of products

divided by number of degrees of freedom) in the same sample, having n degrees of freedom (which ordinarily means that there are $n + 1$ individuals in the sample and that the means are eliminated), is given exactly by the formula³

$$E(\delta s_{ij} \delta s_{km}) = (\sigma_{ik} \sigma_{jm} + \sigma_{im} \sigma_{jk})/n,$$

in which the subscripts may have any values, equal or unequal. When this formula is applied to each of the nine terms of the product and the results are expressed in terms of the correlations ρ_i , there results the asymptotic expression for the covariance given by

$$nE(\delta r_1 \delta r_2) = \frac{1}{2} \rho_1 \rho_2 (\rho_1^2 + \rho_2^2 + \rho_0^2 - 1) + \rho_0 (1 - \rho_1^2 - \rho_2^2).$$

This method provides also one of the derivations of the familiar formula which may be written

$$n\sigma_{r_1}^2 = nE(\delta r_1)^2 = (1 - \rho_1^2)^2, \quad n\sigma_{r_2}^2 = (1 - \rho_2^2)^2.$$

The variance of the difference of r_1 and r_2 is the sum of their variances minus twice their covariance. Hence

$$n\sigma_{r_1-r_2}^2 = (1 - \rho_1^2)^2 + (1 - \rho_2^2)^2 - \rho_1 \rho_2 (\rho_1^2 + \rho_2^2 + \rho_0^2 - 1) + 2\rho_0 (\rho_1^2 + \rho_2^2 - 1).$$

We are testing the hypothesis that $\rho_1 = \rho_2$. If we put a common value ρ for them in the last expression and simplify, we obtain for the standard error of the difference,

$$\sigma_{r_1-r_2} = \sqrt{\frac{(1 - \rho_0)(2 - 3\rho^2 + \rho_0\rho^2)}{n}}.$$

The second factor in parentheses is always positive because of the inequalities limiting the correlations among three variates.

This formula contains two unknown parameters, ρ and ρ_0 . The classical procedure would be substitute r_1 , r_2 and r_0 respectively for ρ_1 , ρ_2 , and ρ_0 in the previous formula, and use the resulting standard error expression as if the ratio to it of $r_1 - r_2$ were normally distributed. A first modification, more in line with modern ideas, would be to use some kind of average of r_1 and r_2 as an estimate of both ρ_1 and ρ_2 , since the null hypothesis tested is that these are equal. But whatever sample estimates we substitute for ρ and ρ_0 , the formula remains unsatisfactory, since no suitable limits of error are available. If instead of the standard error we were to work out the exact distribution of $r_1 - r_2$ we should still not be free from the difficulty. This exact distribution clearly involves both ρ and ρ_0 , since its variance does so. Neither can we escape from the trouble by using some function $z = f(r)$, such as the inverse hyperbolic tangent suggested by R. A. Fisher, and considering the standard error of $z_1 - z_2 =$

³ I have given a derivation of this formula from the characteristic function of the multivariate normal distribution [1]. Numerous special cases appear in earlier literature. The derivation above is a simplification and improvement of several versions, appearing in the various early writings of Karl Pearson.

$f(r_1) - f(r_2)$; for this standard error will have as the first term in its expansion in a series of powers of n^{-1} simply the product of the expression above for $\sigma_{r_1-r_2}$ by $f'(\rho)$; and this must clearly involve both ρ_0 and ρ .

3. Nuisance Parameters. This is not by any means the only statistical problem in which unknown and undesired parameters enter into the distribution of the statistic which we should naturally use to test a hypothesis. Indeed, the early investigation which was perhaps most influential in setting the whole tone of modern statistical research was that [2] in which W. C. Gosset ("Student") arrived at the exact distribution of the ratio of a deviation in the mean to the *estimated* standard error. The previous practice (which unfortunately survives today in some quarters, and is even taught to students without explaining its approximate character) was to neglect the sampling errors in the estimate of the unknown variance σ^2 and to treat the ratio as normally distributed with unit variance. The rigorous derivation by Fisher [3] of the Student distribution makes clear the manner in which the nuisance parameter ϑ may in this, and in some other, problems be eradicated from the distribution through integration, after altering the original statistic (the deviation in the mean) by dividing it by another statistic. The new statistic, the Student ratio, vanishes whenever the old statistic, the deviation in the mean, does so, and the same hypothesis is tested by both. This then is one way to get rid of a nuisance parameter: when you have a statistic estimating a parameter whose vanishing is in question, but whose distribution involves another parameter, alter the statistic by multiplying or dividing by another statistic in such a way that the new function vanishes whenever the old one does so; and *do this in such a way that the new distribution will be independent of the nuisance parameter*. Unhappily, this method has been applied successfully only in particular cases, and no way to use it in the problem at hand has been found.

A second method is that of transformation employed by Fisher in dealing with such problems as testing the significance of the difference between the correlation coefficients in independent samples between the same two variates. The need for the transformation in this case is occasioned by the presence in the distribution of the difference of the sample correlations of the unknown true value, which is not directly relevant to the comparison. We have seen that this method also fails to solve our problem.

A third method of dealing with nuisance parameters is the use of fiducial probability by R. A. Fisher [4] and by Daisy M. Starkey [5] in testing the significance of the difference between the means of two samples when the variances may be unequal. Criticisms of these applications of fiducial probability have been made by M. S. Bartlett [6] and B. L. Welch [7], and the field of applicability of such methods is still in need of elucidation.

Some findings of J. Neyman [8] having a bearing on the general nuisance parameter problem should also be noted.

The only other class of methods for dealing with nuisance parameters of which

I am aware involves the comparison of the particular sample obtained, not with the whole population of samples with which a comparison might be made if we knew the value of the troublesome parameter, but with a sub-population selected with reference to the sample in such a way that the distribution, in this sub-population, of the statistic used does not involve any unknown parameter. An example is the testing of significance of a regression coefficient. Thus if we suppose that a sample of values of x and y is drawn from a bivariate normal population, and calculate the regression coefficient b of y on x in the sample, the distribution of b involves not only the population value β , but also the ratio α of the variances in the population. Since this second parameter is unknown, and can only be estimated from the sample, it is not possible to use the distribution of b in the whole population directly to test the significance of $b - \beta$. What we do is to find the place of this difference, not in the whole population of values in which both x and y are drawn at random, but in a sub-population for which the values of x are the same as in our sample. We may alternatively say that we limit the sub-population only to that for which the sum of the squares of the deviations of the values of x from their mean is the same as in our sample; the results are the same. The distribution in this sub-population of the ratio of $b - \beta$ to its estimated standard error is of the Student form, with no unknown parameters, and on this basis it is possible to make exact and satisfactory tests and to set up fiducial limits for b . Another example is that of contingency tables. The practice now accepted (after a controversy) for testing independence of two modes of classification, such as classification of persons according as they have or have not been vaccinated, and again according as they live through an epidemic or die, is to compare the observed contingency table, not with all possible contingency tables of the same numbers of rows and columns, but only with the possible contingency tables having exactly the same marginal totals as the observed table.

4. An Exact Solution. We shall solve the problem of the significance of the difference of r_1 and r_2 with the understanding that the meaning of significance is to be interpreted by reference to the sub-population of possible samples for which the predictors x_1 and x_2 have the same set of values as those observed in the particular sample available. This procedure, besides yielding an exact distribution without unknown parameters, has the advantage of relaxing the stringency of the requirement of a trivariate normal distribution. We now make only the assumptions customary in the method of least squares, that the predictand y has the univariate normal distribution for each set of values of x_1 and x_2 , independently for the different sets, with a common variance σ^2 , and with the expectation of y for a fixed pair of values of the predictors a linear function of these predictors. No assumption is involved regarding the distribution of the predictors, since we regard them as fixed in all the samples with which we compare our particular sample. The advantages of exactness and of freedom

from the somewhat special trivariate normal assumption are attained at the expense of sacrificing the precise applicability of the results to other sets of values of the predictors.

Since the correlational properties are unchanged by additive and multiplicative constants, we may suppose that

$$(1) \quad Sx_1 = 0 = Sx_2, \quad Sx_1^2 = 1 = Sx_2^2,$$

where S stands for summation over a sample of N individuals. The notation may be made more explicit by the adjunction of an additional subscript α , varying from 1 to N , to denote the individual member of the sample, so that instead of Sx_1 , for example, we might write $Sx_{1\alpha}$. The omission of this additional subscript is convenient and will usually leave no ambiguity when we deal with sums, but it will be convenient to retain it in connection with individual values. The correlation r_0 of x_1 with x_2 in all those samples we shall consider is, by (1)

$$r_0 = Sx_1x_2.$$

Now consider the new quantities

$$(2) \quad x'_\alpha = \frac{x_{1\alpha} - x_{2\alpha}}{\sqrt{2(1 - r_0)}}, \quad x''_\alpha = \frac{x_{1\alpha} + x_{2\alpha}}{\sqrt{2(1 + r_0)}}.$$

Evidently, from (1) and (2),

$$(3) \quad Sx' = 0 = Sx'', \quad Sx'^2 = 1 = Sx''^2, \quad Sx'x'' = 0.$$

Since the mean value $E(y_\alpha)$ is a linear function of $x_{1\alpha}$ and $x_{2\alpha}$, y_α may, upon subtracting a constant from all these expectations, be written

$$(4) \quad y_\alpha = \beta_1x_{1\alpha} + \beta_2x_{2\alpha} + \Delta_\alpha,$$

where $\Delta_1, \dots, \Delta_N$ are normally and independently distributed with variances all equal to σ^2 and expectations zero. The assumption that x_1 and x_2 are equally correlated with y in the population leads to the conclusion that $\beta_1 = \beta_2$; and putting $\beta = \beta_1\sqrt{2(1 + r_0)}$, we then have from (4) and (2):

$$(5) \quad y_\alpha = \beta x''_\alpha + \Delta_\alpha.$$

Consequently, by (3)

$$Sx'y = Sx'_\alpha y_\alpha = \beta Sx'x'' + Sx'\Delta = Sx'\Delta;$$

and this function has a normal distribution with zero mean and variance σ^2 .

If in the sample we work out a regression equation

$$Y = a + b'x' + b''x'',$$

the normal equations for determining b' and b'' must by (3) take the simple forms

$$a = \bar{y}, \quad b' = Sx'y, \quad b'' = Sx''y.$$

From the general theory of least squares it is known that the sum of squares of residuals is

$$Sv^2 = S(y - \bar{y})^2 = Sy^2 - \bar{y}Sy - (Sx'y)^2 - S(x''y)^2,$$

and that Sv^2/σ^2 has the χ^2 distribution with $n = N - 3$ degrees of freedom, independently both of $Sx'y$ and of $Sx''y$. From these facts it follows that

$$(6) \quad t = Sx'y \sqrt{\frac{n}{Sv^2}}$$

has the Student distribution with n degrees of freedom. Since in accordance with the foregoing definitions and (1) we have

$$Sx'y = (r_1 - r_2) \sqrt{\frac{S(y - \bar{y})^2}{2(1 - r_0)}},$$

and since also it is known that

$$Sv^2 = S(y - \bar{y})^2 \frac{D}{1 - r_0^2},$$

where

$$D = \begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_0 \\ r_2 & r_0 & 1 \end{vmatrix},$$

(6) may be written

$$(7) \quad t = (r_1 - r_2) \sqrt{\frac{n(1 + r_0)}{2D}}.$$

The probability of a greater value of $|t|$ is given by tables of the Student distribution with $n = N - 3$. If this probability is sufficiently small (which conventionally means less than .05, or sometimes .01) we have a corresponding degree of confidence that the variate chosen because of a higher correlation in the sample has actually a higher correlation than the other in the population.

5. The Selection of One Variate from Among Three or More. Suppose that we are to choose one of the variates x_1, \dots, x_p in order to predict y . ($p < N - 1$) We choose the one having highest correlation, and wonder how much confidence to place in this choice. We shall now determine the distribution of a function suitable for testing the hypothesis that there is no real difference between any pair of the correlations of x_1, \dots, x_p with y . Again we shall assume the values of these predictors fixed, and look for the place of our particular sample among all samples having these values, with only y free to vary normally by chance.

Let $a_{ij} = S(x_i - \bar{x}_i)(x_j - \bar{x}_j)$, and let c_{ij} be the cofactor of a_{ij} in the determinant a of these quantities, divided by a . Then

$$(8) \quad \sum a_{ij} c_{ik} = \delta_{ik} = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Here Σ stands for summation from 1 to p . Let

$$(9) \quad w_i = \frac{\sum_j c_{ij}}{\sum \sum c_{ij}},$$

$$(10) \quad l_i = S(x_i - \bar{x})y,$$

$$(11) \quad l = \Sigma w_i l_i.$$

From (9) it follows that

$$(12) \quad \Sigma w_i = 1.$$

From the hypothesis that y is in the population equally correlated with all the x_i it follows that l_1, \dots, l_p have equal expectations, which we may denote by λ ; and from (11) and (12) it follows that also $E(l) = \lambda$. Obviously

$$(13) \quad E(l_i - \lambda)(l_j - \lambda) = \sigma^2 a_{ij},$$

where σ^2 is the variance of those values of y corresponding to a fixed set of values of the x 's. From (11), (13) and (9) we obtain

$$(14) \quad E(l - \lambda)^2 = \frac{\sigma^2}{\Sigma \Sigma c_{ij}}.$$

Since the l_i are linear functions of the y 's, they have the multivariate normal distribution. From the theory of this distribution and the values (13) of the covariances it follows that the distribution has the form

$$(2\pi)^{-1/2} a^{-1/2} \sigma^{-p} e^{-T/2\sigma^2} dl_1 \dots dl_p,$$

where a is the determinant of the a_{ij} 's, and

$$T = \Sigma \Sigma c_{ij} (l_i - \lambda)(l_j - \lambda).$$

We may introduce linear functions l'_1, \dots, l'_p of $l_1 - \lambda, \dots, l_p - \lambda$ such that $T = l'^2_1 + \dots + l'^2_p$, and such that $l'^2_p = (l - \lambda)^2 \Sigma \Sigma c_{ij}$. Now $\frac{l'^2_1 + \dots + l'^2_{p-1}}{\sigma^2}$ has the χ^2 distribution with $p - 1$ degrees of freedom. The numerator of this expression equals

$$\begin{aligned} T - l'^2_p &= \Sigma \Sigma c_{ij} (l_i - \lambda)(l_j - \lambda) - (l - \lambda)^2 \Sigma \Sigma c_{ij} \\ &= \Sigma \Sigma c_{ij} l_i l_j - l^2 \Sigma \Sigma c_{ij} \\ &= \Sigma \Sigma c_{ij} (l_i - l)(l_j - l). \end{aligned}$$

The penultimate form shows that this function is independent of λ ; the last, as a positive definite form in the deviations of the l 's from their weighted mean, shows that sufficiently large values of the expression will reveal with definiteness the inequality of the predicting powers of the p variates when this exists.

It is well known that the regression coefficients of y upon the set of variates x_1, \dots, x_p are completely independent of the sum of squares Sv^2 of residuals from the regression equation. Since the l 's are linear functions of these regression coefficients, (namely the linear functions appearing in the normal equations), they also are independent of Sv^2 . Hence, if we put

$$s_1^2 = \frac{\sum \sum c_{ij} l_i l_j - l^2 \sum \sum c_{ij}}{p - 1},$$

$$s_2^2 = \frac{Sv^2}{N - p - 1},$$

the ratio $F = s_1^2/s_2^2$ will, in case of equality of the correlations of the various x 's with y , have the variance ratio distribution with $n_1 = p - 1$ and $n_2 = N - p - 1$ degrees of freedom. When $p = 2$ this test reduces exactly to (7), as it should, and $F = t^2$.

In the numerical application of this method, the regression coefficients b_i of y on x_1, \dots, x_p should first be worked out by the inverse matrix method. The right-hand members of the normal equations are l_1, \dots, l_p , the coefficients in these equations are the a_{ij} , and the calculation of s_1^2 is simplified with the help of the identity

$$\sum \sum c_{ij} l_i l_j = \sum b_i l_i.$$

6. Selection of Additional Variates When Some Have Been Chosen. Suppose now that q predictors have been included definitely in the regression equation, and that one more is to be selected for inclusion among p additional predictors that are available. The criterion now is that that one should be chosen tentatively which has the highest partial correlation with the predictand, eliminating those already definitely chosen; but the confidence to be placed in the choice is to be judged by an adaptation of the criterion of the preceding section. It is only necessary to consider the a_{ij} , l_i , c_{ij} and b_i ($i, j = 1, \dots, p$) as calculated from the new predictors and the deviations of y from the regression equation on the predictors already adopted. Formulae may easily be derived for the values of these quantities in terms of those already found and the sums of products, so as to simplify the calculations. Sv^2 will now stand for the sum of squares of residuals from the regression equation involving all the $p + q$ predictors. It is to be divided by $N - p - q - 1$ to obtain s_2^2 . The numbers of degrees of freedom with respect to which F is to be judged are now $n_1 = p - 1$ and $n_2 = N - p - q - 1$. When $p = 2$ this test, like that of the preceding section, reduces to the use of the t -distribution of (7), with $n = N - q - 3$, and the correlations standing for partial correlations eliminating the predictors already definitely chosen.

A special instance in which this procedure is applicable is in economic time series, in which time, in the form of orthogonal polynomials, must ordinarily be "partialled out" in order that tests of significance may be sound.

7. Further Problems. It is natural to ask whether the foregoing work can be extended to examine the soundness of the selection, on the basis of a greater multiple correlation, of a particular set of two or more variates, chosen from among several such sets. The simplest such problem that goes beyond what has been done above deals with two sets, each of two predictors, having in a sample multiple correlations R and R' with the predictand. The question is whether the difference $R - R'$ is significant.

Suppose that, in the interests of simplicity and the hope of attaining a solution satisfactorily free from unknown parameters, we assume as before that the predictors have a fixed set of values, the same in all samples. Since multiple correlations are invariant under linear transformations of predictors, we may without loss of generality assume that the predictors in each set are mutually uncorrelated and have sums of squares equal to unity. Indeed, we may go somewhat further in standardizing the sets of values to which consideration can be confined without loss of generality, with the help of some ideas introduced in the paper [1]. In the terminology of that paper, the variates in each set may be considered *canonical* with respect to the relationship between the sets. This means that linear functions x_1 and x_2 of the two variates in one set, and linear functions x'_1 and x'_2 of those in the other set, can be chosen so as to satisfy not only the conditions

$$(15) \quad \begin{aligned} Sx_1 &= Sx_2 = Sx'_1 = Sx'_2 = 0 \\ Sx_1^2 &= Sx_2^2 = Sx'^2_1 = Sx'^2_2 = 1 \\ Sx_1x_2 &= 0 = Sx'_1x'_2, \end{aligned}$$

but also the further conditions

$$(16) \quad Sx_1x'_2 = 0 = Sx_2x'_1.$$

This means that, for all the purposes in view, the two sets of predictors can be characterized as to their mutual relationships by the values of the remaining two sums of products, namely

$$c_1 = Sx_1x'_1, \quad c_2 = Sx_2x'_2.$$

In view of the conditions assumed earlier, c_1 and c_2 are what have been called the *canonical correlations* between the two sets.

To the sets thus standardized, the predictand y is related in a manner expressed by the population regression coefficients β_1 and β_2 of y on the first set, and β'_1 and β'_2 on the second. If we take y as having unit variance in the population, the squared multiple correlation coefficients in the two cases will be

$$\rho^2 = \beta_1^2 + \beta_2^2, \quad \rho'^2 = \beta'^2_1 + \beta'^2_2.$$

The hypothesis to be tested is that $\rho = \rho'$. If b_1, b_2, b'_1, b'_2 denote the sample estimates of the regression coefficients, the statistic appropriate for the test would appear necessarily to be proportional to

$$w = \frac{1}{2}(b_1^2 + b_2^2 - b'^2_1 - b'^2_2).$$

The sample regression coefficients are normally distributed, with population correlations equal to the sample correlations among the corresponding predictors. The variance of each is σ^2 . Thus their joint distribution may be written down at once, in a rather simple form in view of (15) and (16). From this it is possible to determine directly the characteristic function $M(t) = Ee^{tw}$ of w . If we write $K(t) = \log M(t)$ we obtain:

$$2K(t) = \Sigma\{(\beta_j^2 - 2c_j\beta_j\beta'_j + \beta_j'^2)t^2 + (\beta_j^2 - \beta_j'^2)t\}\{1 - (1 - c_j^2)t^2\}^{-1} \\ - \Sigma \log \{1 - (1 - c_j^2)t^2\}.$$

Here the summations are with respect to j over the values 1 and 2. If each set of predictors had had s members, the same result would hold for $K(t)$ except that the summations with respect to j would then extend from 1 to s .

This is a very disappointing result because it contains so many parameters. The distribution of w must contain the same parameters as its characteristic function. All the four parameters β_j, β'_j appear in the expression above, though their effective number is reduced to three by the condition that the two sums of squares shall be equal which constitutes the hypothesis under test. The distribution of w thus contains at least three unknown parameters besides σ .

The estimate of variance s^2 obtained from the residuals from the grand regression equation of y on x_1, x_2, x'_1 , and x'_2 is independent of w . Its distribution is of the usual form and involves a parameter, the population variance, which is a function of $\beta_1, \beta_2, \beta'_1$, and β'_2 . We could therefore pass by a single integration from the distribution of w to that of the statistic w/s^2 , which vanishes with w , and which on this account, and on grounds of physical dimensionality, might be considered appropriate to test the hypothesis that $\rho = \rho'$. The question may be raised whether the distribution of this ratio might not be free from parameters. The answer unfortunately is in the negative, as appears from an examination of the characteristic function of the ratio. Even in the simplified case in which all the c_j are equal, a troublesome parameter persists in the distribution.

Thus we meet again the problem of nuisance parameters, and this time no escape is visible. Perhaps some such artifice as those enumerated in paragraph 3 (for example, some further limitation of the sub-population within which we should seek the place of our particular sample) is capable of yielding an exact, or "studentized" distribution; but this has not yet been found. The problem is of considerable interest, not only because of its practical importance, but because of its suggestiveness in connection with general theory.

Numerous other problems having both practical importance and general theoretical interest are associated with the selection of predictors. For example, we have not dealt at all with the problem of the number of predictors that should be used when maximum accuracy in prediction, or in evaluation of the regression coefficients, is the sole criterion. A particular case is the determination of the degree of the regression polynomial which should be fitted to obtain

maximum accuracy, for example of the number of orthogonal polynomials in fitting a trend. Such customary criteria as minimizing the *estimated* variance of deviations, in which the sum of squares which is the numerator and the number of degrees of freedom which is the denominator both diminish to zero as the number of variates is increased, do not rest upon any satisfactory general theory.

Another related set of problems is concerned with variates more numerous than the observations on each. It is clear that there is real information inherent in data of this kind, but existing theory and methods, including those of the present paper, are not adequate to utilize it in a thoroughly efficient manner. A recent paper of P. L. Hsu [9] is unique in not excluding the case in which the variates outnumber the observations.

8. Summary. A criterion has been obtained for judging the definiteness of the selection of a particular variate, from among several available for prediction, on the basis of its having the maximum sample correlation with the predictand. A variation of this criterion is applied in paragraph 6 to the problem of extending the list of variates to be used in a regression formula.

Some of the problems of "nuisance parameters" which affect general theory are illustrated in this problem. Some outstanding unsolved problems related to these questions are discussed in paragraph 7.

REFERENCES

- [1] Harold Hotelling, "Relations Between Two Sets of Variates," *Biometrika*, Vol. 28 (1936), pp. 321-377.
- [2] "Student," "The Probable Error of a Mean," *Biometrika*, Vol. 6 (1908), pp. 1-25.
- [3] R. A. Fisher, "Applications of 'Student's' Distribution," *Metron*, Vol. 5 (1925), pp. 90-104.
- [4] R. A. Fisher, "The Fiducial Argument in Statistical Inference," *Annals of Eugenics*, Vol. 6 (1935), pp. 391-398. See also Fisher's answer to Bartlett in the *Annals of Math. Stat.*, Vol. 10 (1939), pp. 383-388 and the references there given.
- [5] Daisy M. Starkey, "A Test of the Significance of the Difference Between Means of Samples from Two Normal Populations Without Assuming Equal Variances," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 201-213.
- [6] M. S. Bartlett, "The Information Available in Small Samples," *Proc. Camb. Phil. Soc.*, Vol. 32 (1936), pp. 560-566.
- [7] B. L. Welch, "On Confidence Limits and Sufficiency, with Particular Reference to Parameters of Location," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 58-69.
- [8] *Statistical Research Memoirs*, Vol. 2 (1938), pp. 58-59.
- [9] P. L. Hsu, "On the Distribution of Roots of Certain Determinantal Equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 250-258.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

THE FITTING OF STRAIGHT LINES IF BOTH VARIABLES ARE SUBJECT TO ERROR

BY ABRAHAM WALD

1. Introduction. The problem of fitting straight lines if both variables x and y are subject to error, has been treated by many authors. If we have $N > 2$ observed points (x_i, y_i) ($i = 1, \dots, N$), the usually employed method of least squares for determining the coefficients a, b , of the straight line $y = ax + b$ is that of choosing values of a and b which minimize the sum of the squares of the residuals of the y 's, i.e. $\sum (ax_i + b - y_i)^2$ is a minimum. It is well known that treating y as an independent variable and minimizing the sum of the squares of the residuals of the x 's, we get a different straight line as best fit. It has been pointed out¹ that if both variables are subject to error there is no reason to prefer one of the regression lines described above to the other. For obtaining the "best fit," which is not necessarily equal to one of the two lines mentioned, new criteria have to be found. This problem was treated by R. J. Adcock as early as 1877.²

He defines the line of best fit as the one for which the sum of the squares of the normal deviates of the N observed points from the line becomes a minimum. (Another early attempt to solve this problem by minimizing the sum of squares of the normal deviates was made by Karl Pearson.³)

Many objections can be raised against this method. First, there is no justification for minimizing the sum of the squares of the *normal* deviates, and not the deviations in some other direction. Second, the straight line obtained by that method is not invariant under transformation of the coordinate system. It is clear that a satisfactory method should give results which do not depend on the choice of a particular coordinate system. This point has been emphasized by C. F. Roos. He gives⁴ a good summary of the different methods and then proposes a general formula for fitting lines (and planes in case of more than two variables) which do not depend on the choice of the coordinate system.

¹ See for instance Henry Schultz' "The Statistical Law of Demand," *Jour. of Political Economy*, Vol. 33, Dec. (1925).

² *Analyst*, Vol. IV, p. 183 and Vol. V, p. 53.

³ "On Lines and Planes of Closest Fit to Systems of Points in Space" *Phil. Mag.* 6th Ser. Vol. II (1901).

⁴ "A General Invariant Criterion of Fit for Lines and Planes where all Variates are Subject to Error," *Metron*, February 1937. See also Oppenheim and Roos *Bulletin of the American Mathematical Society*, Vol. 34 (1928), pp. 140-141.

Roos' formula includes many previous solutions⁵ as special cases. H. E. Jones⁶ gives an interesting geometric interpretation of Roos' general formula.

It is a common feature of Roos' general formula and of all other methods proposed in recent years that the fitted straight line cannot be determined without *a priori* assumptions (independent of the observations) regarding the weights of the errors in the variables x and y . That is to say, either the standard deviations of the errors in x and in y are involved (or at least their ratio is included) in the formula of the fitted straight line and there is no method given by which those standard deviations can be estimated by means of the observed values of x and y .

R. Frisch⁷ has developed a new general theory of linear regression analysis, when all variables are subject to error. His very interesting theory employs quite new methods and is not based on probability concepts. Also on the basis of Frisch's discussion it seems that there is no way of determining the "true" regression without *a priori* assumptions about the disturbing intensities.

T. Koopmans⁸ combined Frisch's regression theory with the classical one in a new general theory based on probability concepts. Also, according to his theory, the regression line can be determined only if the ratio of the standard deviations of the errors is known.

In a recent paper R. G. D. Allen⁹ gives a new interesting method for determining the fitted straight line in case of two variables x and y . Denoting by σ_x the standard deviation of the errors in x , by σ_y the standard deviation of the errors in y and by ρ the correlation coefficient between the errors in the two variables, Allen emphasizes (p. 194)⁹ that the fitted line can be determined only if the values of two of the three quantities σ_x , σ_y , ρ are given *a priori*.

Finally I should like to mention a paper by C. Eisenhart,¹⁰ which contains many interesting remarks related to the subject treated here.

In the present paper I shall deal with the case of two variables x and y in which the errors are uncorrelated. It will be shown that under certain conditions:

- (1) The fitted straight line can be determined without making *a priori* assumptions (independent of the observed values x and y) regarding the standard deviations of the errors.
- (2) The standard deviation of the errors can be well estimated by means of

⁵ For instance also Corrado Gini's method described in his paper, "Sull' Interpolazione di una Retta Quando i Valori della Variable Indipendente sono Affetti da Errori Accidentali," *Metron*, Vol. I, No. 3 (1921), pp. 63-82.

⁶ "Some Geometrical Considerations in the General Theory of Fitting Lines and Planes," *Metron*, February 1937.

⁷ *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo, 1934.

⁸ *Linear Regression Analysis of Economic Time Series*, Haarlem, 1937.

⁹ "The Assumptions of Linear Regression," *Economica*, May 1939.

¹⁰ "The interpretation of certain regression methods and their use in biological and industrial research," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 162-186.

the observed values of x and y . The precision of the estimate increases with the number of the observations and would give the exact values if the number of observations were infinite. (See in this connection also condition V in section 3.)

2. Formulation of the Problem. Let us begin with a precise formulation of the problem. We consider two sets of random variables¹¹

$$x_1, \dots, x_N; \quad y_1, \dots, y_N.$$

Denote the expected value $E(x_i)$ of x_i by X_i and the expected value $E(y_i)$ of y_i by Y_i ($i = 1, \dots, N$). We shall call X_i the true value of x_i , Y_i the true value of y_i , $x_i - X_i = \epsilon_i$ the error in the i -th term of the x -set, and $y_i - Y_i = \eta_i$ the error in the i -th term of the y -set.

The following assumptions will be made:

I. The random variables $\epsilon_1, \dots, \epsilon_N$ each have the same distribution and they are uncorrelated, i.e. $E(\epsilon_i \epsilon_j) = 0$ for $i \neq j$. The variance of ϵ_i is finite.

II. The random variables η_1, \dots, η_N each have the same distribution and are uncorrelated, i.e. $E(\eta_i \eta_j) = 0$ for $i \neq j$. The variance of η_i is finite.

III. The random variables ϵ_i and η_j ($i = 1, \dots, N; j = 1, \dots, N$) are uncorrelated, i.e. $E(\epsilon_i \eta_j) = 0$.

IV. A single linear relation holds between the true values X and Y , that is to say $Y_i = \alpha X_i + \beta$ ($i = 1, \dots, N$).

Denote by ϵ a random variable having the same probability distribution as possessed by each of the random variables $\epsilon_1, \dots, \epsilon_N$, and by η a random variable having the same distribution as η_1, \dots, η_N .

The problem to be solved can be formulated as follows:

We know only two sets of observations: $x'_1, \dots, x'_N; y'_1, \dots, y'_N$, where x'_i denotes the observed value of x_i and y'_i denotes the observed value of y_i . We know neither the true values $X_1, \dots, X_N; Y_1, \dots, Y_N$, nor the coefficients α and β of the linear relation between them. We have to estimate by means of the observations $x'_1, \dots, x'_N; y'_1, \dots, y'_N$, (1) the values of α and β , (2) the standard deviation σ_ϵ of ϵ , and (3) the standard deviation σ_η of η .

Problems of this kind occur often in Economics, where we are dealing with time series. For example, denote by x_i the price of a certain good G in the period t_i , and by y_i the quantity of G demanded in t_i . In each time period t_i there exists a normal price X_i and a normal demand Y_i which would obtain if the influence of some accidental disturbances could be eliminated. If we have reason to assume that there exists between the normal price and the normal demand a linear relationship we have to deal with a problem of the kind described above.

In the following discussions we shall use the notations x_i and y_i also for their

¹¹ A random or stochastic variable is a real variable associated with a probability distribution.

observed values x'_i and y'_i since it will be clear in which sense they are meant and no confusion can arise.

3. Consistent Estimates of the Parameters α , β , σ_ϵ , σ_η . For the sake of simplicity we assume that N is even. We consider the expression

$$(1) \quad \begin{aligned} a_1 &= \frac{(x_1 + \dots + x_m) - (x_{m+1} + \dots + x_N)}{N}, \\ a_2 &= \frac{(y_1 + \dots + y_m) - (y_{m+1} + \dots + y_N)}{N}, \end{aligned}$$

where $m = N/2$. As an estimate of α we shall use the expression

$$(2) \quad a = \frac{a_2}{a_1} = \frac{(y_1 + \dots + y_m) - (y_{m+1} + \dots + y_N)}{(x_1 + \dots + x_m) - (x_{m+1} + \dots + x_N)}.$$

We make the assumption

V. *The limit inferior of*

$$\left| \frac{(X_1 + \dots + X_m) - (X_{m+1} + \dots + X_N)}{N} \right| \quad (N = 2, 3, \dots \text{ad. inf.})$$

is positive.

We shall prove that a is a consistent estimate of α , i.e. a converges stochastically to α with $N \rightarrow \infty$, if the assumptions I-V hold. Denote the expected value of a_1 by \bar{a}_1 and the expected value of a_2 by \bar{a}_2 . It is obvious that

$$(3) \quad \begin{aligned} \bar{a}_1 &= \frac{(X_1 + \dots + X_m) - (X_{m+1} + \dots + X_N)}{N}, \\ \bar{a}_2 &= \frac{(Y_1 + \dots + Y_m) - (Y_{m+1} + \dots + Y_N)}{N}. \end{aligned}$$

On account of the condition IV we have

$$(4) \quad \bar{a}_2 = \alpha \bar{a}_1, \quad \text{or} \quad \frac{\bar{a}_2}{\bar{a}_1} = \alpha.$$

The variance of $a_1 - \bar{a}_1$ is equal to σ_ϵ^2/N and the variance of $a_2 - \bar{a}_2$ is equal to σ_η^2/N . Hence a_1 and a_2 converge stochastically towards \bar{a}_1 and \bar{a}_2 respectively.

From that and assumption V it follows that also $\frac{a_2}{a_1}$ converges stochastically towards $\frac{\bar{a}_2}{\bar{a}_1} = \alpha$. The intercept β of the regression line will be estimated by

$$(5) \quad b = \bar{y} - a\bar{x}, \quad \text{where } \bar{x} = \frac{x_1 + \dots + x_N}{N} \quad \text{and} \quad \bar{y} = \frac{y_1 + \dots + y_N}{N}.$$

Denote by \bar{X} the arithmetic mean of X_1, \dots, X_N and by \bar{Y} the arithmetic mean of Y_1, \dots, Y_N . Since \bar{y} converges stochastically towards \bar{Y} , \bar{x} towards

\bar{X} , and a towards α , b converges stochastically towards $\bar{Y} - \alpha\bar{X}$. From condition IV it follows that $\bar{Y} - \alpha\bar{X} = \beta$. Hence b converges stochastically towards β .

Let us introduce the following notations:

$$s_x = \sqrt{\sum \frac{(x_i - \bar{x})^2}{N}} = \text{sample standard deviation of the } x\text{-observations,}$$

$$s_y = \sqrt{\sum \frac{(y_i - \bar{y})^2}{N}} = \text{sample standard deviation of the } y\text{-observations,}$$

$$s_{xy} = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{N} = \text{sample covariance between the } x\text{-set and } y\text{-set.}$$

s_x , s_y and s_{xy} denote the same expressions of the true values X_1, \dots, X_N ; Y_1, \dots, Y_N .

It is obvious that

$$(6) \quad E(s_x^2) = s_x^2 + \sigma_\epsilon^2 \frac{N-1}{N},$$

$$(7) \quad E(s_y^2) = s_y^2 + \sigma_\eta^2 \frac{N-1}{N},$$

$$(8) \quad E(s_{xy}) = s_{xy},$$

where $E(s_x^2)$, $E(s_y^2)$, and $E(s_{xy})$ denote the expected values of s_x^2 , s_y^2 , and s_{xy} .¹²

Since $Y_i = \alpha X_i + \beta$, we have

$$(9) \quad s_y = \alpha s_x,$$

$$(10) \quad s_{xy} = \alpha s_x^2.$$

From (8), (9) and (10) we get

$$(11) \quad s_x^2 = \frac{E(s_{xy})}{\alpha},$$

$$(12) \quad s_y^2 = \alpha E(s_{xy}).$$

If we substitute in (6) and (7) for s_x^2 and s_y^2 their values in (11) and (12), we get

$$(13) \quad \sigma_\epsilon^2 = \left[E(s_x^2) - \frac{E(s_{xy})}{\alpha} \right] N / (N-1),$$

$$(14) \quad \sigma_\eta^2 = [E(s_y^2) - \alpha E(s_{xy})] N / (N-1).$$

¹² I observe that the equations (6), (7) and (8) are essentially the same as those investigated by R. Frisch, *Statistical Confluence Analysis* pp. 51-52. See also Allen's equations (4) l.c. p. 194.

Since s_x^2 , s_y^2 , s_{xy} converge stochastically towards their expected values and a converges stochastically towards α , the expressions

$$(15) \quad \left[s_x^2 - \frac{s_{xy}^2}{a} \right] N / (N - 1)$$

and

$$(16) \quad [s_y^2 - a s_{xy}] N / (N - 1)$$

are consistent estimates of σ_ϵ^2 and σ_η^2 respectively.

4. Confidence Interval for α . In this section, as well as in sections 5 and 6, only the assumptions I-IV are assumed to hold. In other words, all statements made in these sections are valid independently of Assumption V, except where the contrary is explicitly stated.

Let us introduce the following notation:

$$\begin{aligned} \bar{x}_1 &= \frac{x_1 + \dots + x_m}{m}; & \bar{y}_1 &= \frac{y_1 + \dots + y_m}{m} \\ \bar{x}_2 &= \frac{x_{m+1} + \dots + x_N}{m}; & \bar{y}_2 &= \frac{y_{m+1} + \dots + y_N}{m} \\ (s'_x)^2 &= \frac{\sum_{i=1}^m (x_i - \bar{x}_1)^2 + \sum_{j=m+1}^N (x_j - \bar{x}_2)^2}{N} \\ (s'_y)^2 &= \frac{\sum_{i=1}^m (y_i - \bar{y}_1)^2 + \sum_{j=m+1}^N (y_j - \bar{y}_2)^2}{N} \\ s'_{xy} &= \frac{\sum_{i=1}^m (x_i - \bar{x}_1)(y_i - \bar{y}_1) + \sum_{j=m+1}^N (x_j - \bar{x}_2)(y_j - \bar{y}_2)}{N} \end{aligned}$$

\bar{X}_1 , \bar{X}_2 , \bar{Y}_1 , \bar{Y}_2 , $(s'_x)^2$, $(s'_y)^2$ and s'_{xy} denote the same functions of the true values X_1, \dots, X_N , Y_1, \dots, Y_N . The expressions s'_x , s'_y , and s'_{xy} are slightly different from the corresponding expressions s_x , s_y , and s_{xy} . The reason for introducing these new expressions is that the distributions of s_x , s_y , and s_{xy} are not independent of the slope $a = \frac{a_2}{a_1}$ of the sample regression line, but s'_x , s'_y and s'_{xy} are distributed independently from a (assuming that ϵ and η are normally distributed). The latter statement follows easily from the fact that according to (1) and (2) $a = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2}$ and s'_x , s'_y , s'_{xy} are distributed independently of \bar{x}_1 , \bar{x}_2 , \bar{y}_1 and \bar{y}_2 .

In the same way as we derived (13) and (14), we get

$$(13') \quad \sigma_{\epsilon}^2 = \left[E(s'_x)^2 - \frac{E(s'_{xy})}{\alpha} \right] N / (N - 2),$$

$$(14') \quad \sigma_{\eta}^2 = [E(s'_y)^2 - \alpha E(s'_{xy})] N / (N - 2).$$

These formulae differ from the corresponding formulae (13) and (14) only in the denominator of the second factor, having there $N - 2$ instead of $N - 1$. This is due to the fact that the estimates s_x , s_y , s_{xy} are based on $N - 1$ degrees of freedom whereas s'_x , s'_y and s'_{xy} are based only on $N - 2$ degrees of freedom. From (13') and (14') we get the following estimates¹³ for σ_{ϵ}^2 and σ_{η}^2 :

$$(17) \quad \left[(s'_x)^2 - \frac{s'_{xy}}{\alpha} \right] N / (N - 2),$$

$$(18) \quad [(s'_y)^2 - \alpha s'_{xy}] N / (N - 2).$$

Hence we get as an estimate of $\sigma_{\eta}^2 + \alpha^2 \sigma_{\epsilon}^2$ the expression:

$$(19) \quad s^2 = [(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}] N / (N - 2) \\ = \frac{N}{N - 2} \left\{ \frac{\sum_{i=1}^m [(y_i - \alpha x_i) - (\bar{y}_1 - \alpha \bar{x}_1)]^2 + \sum_{j=m+1}^N [(y_j - \alpha x_j) - (\bar{y}_2 - \alpha \bar{x}_2)]^2}{N} \right\}$$

Now we shall show that

$$(20) \quad \frac{(N - 2)s^2}{\sigma_{\eta}^2 + \alpha^2 \sigma_{\epsilon}^2}$$

has the χ^2 -distribution with $N - 2$ degrees of freedom, provided that ϵ and η are normally distributed. In fact,

$$(y_i - \alpha x_i) - (\bar{y}_1 - \alpha \bar{x}_1) = \eta_i - \alpha \epsilon_i - (\eta_1 - \alpha \bar{\epsilon}_1) \quad (i = 1, \dots, m)$$

and

$$(y_j - \alpha x_j) - (\bar{y}_2 - \alpha \bar{x}_2) = \eta_j - \alpha \epsilon_j - (\eta_2 - \alpha \bar{\epsilon}_2) \quad (j = m + 1, \dots, N),$$

where

$$\bar{\epsilon}_1 = \frac{\epsilon_1 + \dots + \epsilon_m}{m}, \quad \bar{\epsilon}_2 = \frac{\epsilon_{m+1} + \dots + \epsilon_N}{m}, \\ \eta_1 = \frac{\eta_1 + \dots + \eta_m}{m}, \quad \eta_2 = \frac{\eta_{m+1} + \dots + \eta_N}{m}$$

Since the variance of $\eta_k - \alpha \epsilon_k$ is equal to $\sigma_{\eta}^2 + \alpha^2 \sigma_{\epsilon}^2$ and since $\eta_k - \alpha \epsilon_k$ is uncorrelated with $\eta_l - \alpha \epsilon_l$ ($k \neq l$) ($k, l = 1, \dots, N$), the expression (20) has the χ^2 -distribution with $N - 2$ degrees of freedom.

¹³ An "estimate" is usually a function of the observations not involving any unknown parameters. We designate here as estimates also some functions involving the parameter α .

Now we shall show that

$$(21) \quad \frac{\sqrt{N} a_1(a - \alpha)}{\sqrt{\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2}}$$

is normally distributed with zero mean and unit variance. In fact from the equations (1)–(4) it follows that

$$\begin{aligned} a_1(a - \alpha) &= \bar{a}_2 + \frac{\bar{\eta}_1 - \bar{\eta}_2}{2} - a_1 \left(\frac{\bar{a}_2}{\bar{a}_1} \right) \\ &= \bar{a}_2 + \frac{\bar{\eta}_1 - \bar{\eta}_2}{2} - \left(\bar{a}_1 + \frac{\bar{\epsilon}_1 - \bar{\epsilon}_2}{2} \right) \left(\frac{\bar{a}_2}{\bar{a}_1} \right) \\ &= \frac{\bar{\eta}_1 - \bar{\eta}_2}{2} - \alpha \frac{\bar{\epsilon}_1 - \bar{\epsilon}_2}{2}. \end{aligned}$$

Since the latter expression is normally distributed (provided that ϵ and η are normally distributed) with zero mean and variance $\frac{\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2}{N}$, our statement about (21) is proved.

Obviously (20) and (21) are independently distributed, hence $\sqrt{N-2}$ times the ratio of (21) to the square root of (20), namely,

$$(22) \quad t = \sqrt{N-2} \frac{\sqrt{N} a_1(a - \alpha)}{\sqrt{N-2} s} = \frac{a_1(a - \alpha) \sqrt{N-2}}{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}}}$$

has the Student distribution with $N-2$ degrees of freedom. Denote by t_0 the critical value of t corresponding to a chosen probability level. The deviation of a from an assumed population value α is significant if

$$\left| \frac{a_1(a - \alpha) \sqrt{N-2}}{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}}} \right| \geq t_0.$$

The confidence interval for α can be obtained by solving the equation in α ,

$$(23) \quad a_1^2(a - \alpha)^2 = [(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}] \frac{t_0^2}{N-2}.$$

Now we shall show that if the relation

$$(24) \quad a_1^2 > \frac{(s'_x)^2 t_0^2}{N-2},$$

holds, the roots α_1 and α_2 are real and a is contained in the interior of the interval $[\alpha_1, \alpha_2]$. From (19) it follows that

$$(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy} > 0$$

for all values of α . Hence, for $\alpha = a$ the left hand side of (23) is smaller than the right hand side. On account of (24) there exists a value $a' > a$ and a

value $a'' < a$ such that the left hand side of (23) is greater than the right hand side for $\alpha = a'$ and $\alpha = a''$. Hence one root must lie between a and a' and the other root between a'' and a . This proves our statement. The relation (24) always holds for sufficiently large N if Assumption V is fulfilled. The confidence interval of α is the interval $[\alpha_1, \alpha_2]$. For very small N (24) may not hold.

Finally I should like to remark that no essentially better estimate of the variance of $\eta - \alpha\epsilon$ can be given than the expression s^2 in (19). In fact, we have $2N$ observations $x_1, \dots, x_N; y_1, \dots, y_N$. For the estimation of the variance of $\eta - \alpha\epsilon$ we must eliminate the unknowns X_1, \dots, X_N and β . (The unknowns Y_1, \dots, Y_N are determined by the relations $Y_i = \alpha X_i + \beta$ and α is involved in the expression whose variance is to be determined.) Hence we have at most $N - 1$ degrees of freedom and the estimate in (19) is based on $N - 2$ degrees of freedom.

5. Confidence Interval for β if α is Given. In this case the best estimate of β is given by the expression:

$$b_\alpha = \bar{y} - \alpha \bar{x} \text{ where } \bar{x} = \frac{x_1 + \dots + x_N}{N} \text{ and } \bar{y} = \frac{y_1 + \dots + y_N}{N}.$$

We have

$$b_\alpha - \beta = (\bar{y} - \bar{Y}) - \alpha(\bar{x} - \bar{X}) = \eta - \alpha\bar{\epsilon}$$

where

$$\bar{\epsilon} = \frac{\epsilon_1 + \dots + \epsilon_N}{N}, \text{ and } \eta = \frac{\eta_1 + \dots + \eta_N}{N}.$$

Hence,

$$(25) \quad \frac{\sqrt{N} (b_\alpha - \beta)}{\sqrt{\sigma_\eta^2 + \alpha^2 \sigma_\epsilon^2}}$$

is normally distributed with zero mean and unit variance. It is obvious that the expressions (20) and (25) are independently distributed. Hence $\sqrt{N - 2}$ times the ratio of (25) to the square root of (20), i.e.

$$t = \sqrt{N - 2} \frac{\sqrt{N} (b_\alpha - \beta)}{\sqrt{N - 2} s} = \frac{\sqrt{N - 2} (b_\alpha - \beta)}{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}}}$$

has the Student distribution with $N - 2$ degrees of freedom. Denoting by t_0 the critical value of t according to the chosen probability level, the confidence interval for β is given by the interval:

$$\left[b_\alpha + \frac{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}}}{\sqrt{N - 2}} t_0, \quad b_\alpha - \frac{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}}}{\sqrt{N - 2}} t_0 \right].$$

6. Confidence Region for α and β Jointly. In most practical cases we want to know confidence limits for α and β jointly. A pair of values α, β can be represented in the plane by the point with the coordinates α, β . A region R of this plane is called confidence region of the true point (α, β) corresponding to the probability level P if the following two conditions are fulfilled.

(1) The region R is a function of the observations $x_1, \dots, x_N; y_1, \dots, y_N$, i.e. it is uniquely determined by the observations.

(2) Before performing the experiment the probability that we shall obtain observed values such that (α, β) will be contained in R , is exactly equal to P . P is usually chosen to be equal to .95 or .99.

We have shown that the expressions (21) and (25), i.e.

$$\frac{\sqrt{N} a_1(a - \alpha)}{\sqrt{\sigma_y^2 + \alpha^2 \sigma_x^2}}, \quad \frac{\sqrt{N} (b_a - \beta)}{\sqrt{\sigma_y^2 + \alpha^2 \sigma_x^2}}$$

are normally distributed with zero mean and unit variance. Now we shall show that these two quantities are independently distributed. For this purpose we have only to show that \bar{x}, \bar{y}, a_1 and a_2 are independently distributed (a_1 and a_2 are defined in (1)), but since

$$a_1 - E(a_1) = (\bar{\epsilon}_1 - \bar{\epsilon}_2)/2$$

$$a_2 - E(a_2) = (\eta_1 - \eta_2)/2$$

$$\bar{x} - E(\bar{x}) = \bar{\epsilon}$$

$$\bar{y} - E(\bar{y}) = \eta,$$

we have only to show that $\bar{\epsilon}, \eta, \bar{\epsilon}_1 - \bar{\epsilon}_2, \eta_1 - \eta_2$ are independently distributed. We obviously have

$$\bar{\epsilon} = \frac{\bar{\epsilon}_1 + \bar{\epsilon}_2}{2}, \quad \eta = \frac{\eta_1 + \eta_2}{2}.$$

It is evident that $\bar{\epsilon}_1, \bar{\epsilon}_2, \eta_1$ and η_2 are independently distributed. Hence, $E[\bar{\epsilon}(\bar{\epsilon}_1 - \bar{\epsilon}_2)] = (E\bar{\epsilon}_1^2 - E\bar{\epsilon}_2^2)/2 = 0$ and also $E[\eta(\eta_1 - \eta_2)] = (E\eta_1^2 - E\eta_2^2)/2 = 0$. Since $\bar{\epsilon}_1 - \bar{\epsilon}_2, \eta_1 - \eta_2$, and $\bar{\epsilon}$ and η are normally distributed, the independence of this set of variables is proved, and therefore also (21) and (25) are independently distributed. It is obvious that the expression (20) is distributed independently of (21) and (25). From this it follows that

$$(26) \quad \frac{N-2}{2} \cdot \frac{N[a_1^2(a-\alpha)^2 + (\bar{y} - \alpha\bar{x} - \beta)^2]}{(N-2)s^2} \\ = \frac{(N-2)[a_1^2(a-\alpha)^2 + (\bar{y} - \alpha\bar{x} - \beta)^2]}{2[(s'_y)^2 + \alpha^2(s'_x)^2 - 2\alpha s'_{xy}]}$$

has the F -distribution (analysis of variance distribution) with 2 and $N-2$ degrees of freedom. The F -distribution is tabulated in Snedecor's book: *Calcu-*

lation and Interpretation of Analysis of Variance, Collegiate Press, Ames, Iowa, 1934. The distribution of $\frac{1}{2} \log F = z$ is tabulated in R. A. Fisher's book: *Statistical Methods for Research Workers*, London, 1936. Denote by F_0 the critical value of F corresponding to the chosen probability level P . Then the confidence region R is the set of points (α, β) which satisfy the inequality

$$(27) \quad \frac{N-2}{2} \cdot \frac{a_1^2(a-\alpha)^2 + (\bar{y} - \alpha\bar{x} - \beta)^2}{(s_y')^2 + \alpha^2(s_x')^2 - 2\alpha s_{xy}'} < F_0.$$

The boundary of the region is given by the equation

$$(28) \quad a_1^2(a-\alpha)^2 + (\bar{y} - \alpha\bar{x} - \beta)^2 = \frac{2F_0}{N-2} [(s_y')^2 + \alpha^2(s_x')^2 - 2\alpha s_{xy}'].$$

This is the equation of an ellipse. Hence the region R is the interior of the ellipse defined by the equation (28). If Assumption V holds, the length of the axes of the ellipse are of the order $1/\sqrt{N}$, hence with increasing N the ellipse reduces to a point.

7. The Grouping of the Observations. We have divided the observations in two equal groups G_1 and G_2 , G_1 containing the first half $(x_1, y_1), \dots, (x_m, y_m)$ and G_2 the second half $(x_{m+1}, y_{m+1}), \dots, (x_N, y_N)$ of the observations. All the formulas and statements of the previous sections remain exactly valid for any arbitrary subdivision of the observations in two equal groups, provided that the subdivision is defined independently of the errors $\epsilon_1, \dots, \epsilon_N$; η_1, \dots, η_N . The question of which is the most advantageous grouping arises, i.e. for which grouping will a be the most efficient estimate of α (will lead to the shortest confidence interval for α). It is easy to see that the greater $|a_1|$ the more efficient is the estimate a of α . The expression $|a_1|$ becomes a maximum if we order the observations such that $x_1 \leq x_2 \leq \dots \leq x_N$. That is to say $|a_1|$ becomes a maximum if we group the observations according to the following:

RULE I. The point (x_i, y_i) belongs to the group G_1 if the number of elements x_j ($j \neq i$) of the series x_1, \dots, x_N for which $x_j \leq x_i$ is less than $m = N/2$. The point (x_i, y_i) belongs to G_2 if the number of elements x_j ($j \neq i$) for which $x_j \leq x_i$ is greater than or equal to m .

This grouping, however, depends on the observed values x_1, \dots, x_N and is therefore in general not entirely independent of the errors $\epsilon_1, \dots, \epsilon_N$. Let us now consider the grouping according to the following:

RULE II. The point (x_i, y_i) belongs to the group G_1 if the number of elements X_j of the series X_1, \dots, X_N for which $X_j \leq X_i$ ($j \neq i$) is less than m . The point (x_i, y_i) belongs to G_2 if the number of elements X_j for which $X_j \leq X_i$ ($j \neq i$) is equal to or greater than m .

The grouping according to Rule II is entirely independent of the errors $\epsilon_1, \dots, \epsilon_N; \eta_1, \dots, \eta_N$. It is identical with the grouping according to Rule I in the following case: Denote by x the median of x_1, \dots, x_N ; assume that ϵ can take values only within the finite interval $[-c, +c]$ and that all the values x_1, \dots, x_N fall outside the interval $[x - c, x + c]$. It is easy to see that in this case $x_i \leq x$ ($i = 1, \dots, N$) holds if and only if $X_i \leq X$, where X denotes the median of X_1, \dots, X_N . Hence the grouping according to Rule II is identical to that according to Rule I and therefore the grouping according to Rule I is independent of the errors $\epsilon_1, \dots, \epsilon_N$. In such cases we get the best estimate of α by grouping the observations according to Rule I. Practically, we can use the grouping according to Rule I and regard it as independent of the errors $\epsilon_1, \dots, \epsilon_N; \eta_1, \dots, \eta_N$ if there exists a positive value c for which the probability that $|\epsilon| \geq c$ is negligibly small and the number of observations contained in $[x - c, x + c]$ is also very small.

Denote by a' the value of a which we obtain by grouping the observations according to Rule I and by a'' the value of a if we group the observations according to Rule II. The value a'' is in general unknown, since the values X_1, \dots, X_N are unknown, except in the special case considered above, when we have $a'' = a'$. We will now show that an upper and a lower limit for a'' can always be given. First, we have to determine a positive value c such that the probability that $|\epsilon| \geq c$ is negligibly small. The value of c may often be determined before we make the observations having some *a priori* knowledge about the possible range of the errors. If this is not the case, we can estimate the value of c from the data. It is well known that if we have errors in both variables and fit a straight line by the method of least squares minimizing in the x -direction, the sum of the squared deviations divided by the number of degrees of freedom will overestimate σ_ϵ^2 . Hence, if ϵ is normally distributed, we can consider the interval $[-3v, 3v]$ as the possible range of ϵ , i.e. $c = 3v$, where v^2 denotes the sum of the squared residuals divided by the number of degrees of freedom. If the distribution of ϵ is unknown, we shall have to take for c a somewhat larger value, for instance $c = 5v$. After having determined c , upper and lower limits for a'' can be given as follows: we consider the system S of all possible groupings satisfying the conditions:

- (1) If $x_i \leq x - c$ the point (x_i, y_i) belongs to the group G_1 .
- (2) If $x_i \geq x + c$ the point (x_i, y_i) belongs to the group G_2 .

We calculate the value of a according to each grouping of the system S and denote the minimum of these values by a^* , and the maximum by a^{**} . Since the grouping according to Rule II is contained in the system S , a^* is a lower and a^{**} an upper limit of a'' .

Let g be a grouping contained in S and denote by I_g the confidence interval for α which we obtain from formula (23) using the grouping g . Denote further by I the smallest interval which contains the intervals I_g for all elements g of S . Then I contains also the confidence interval corresponding to the grouping according to Rule II. If we denote by P the chosen probability level (say

$P = .95$), then we can say: If we were to draw a sample consisting of N pairs of observations $(x_1, y_1), \dots, (x_N, y_N)$, the probability is greater than or equal to P that we shall obtain a system of observations such that the interval I will include the true slope α .

The computing work for the determination of I may be considerable if the number of observations within the interval $[x - c, x + c]$ is not small. We can get a good approximation to I by less computation work as follows: First we calculate the slope a' using the grouping according to Rule I and determine the confidence interval $[a' - \delta, a' + \Delta]$ according to formula (23). Denote by $a(g)$ the value of the slope, i.e. the value of $\frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2}$, corresponding to a grouping g of the system S , and by $[a(g) - \delta_g, a(g) + \Delta_g]$ the corresponding confidence interval calculated from (23). Neglecting the differences $(\delta_g - \delta)$ and $(\Delta_g - \Delta)$, we obtain for I the interval $[a^* - \delta, a^{**} + \Delta]$.

If the difference $a^{**} - a^*$ is small, we can consider $I = [a^* - \delta, a^{**} + \Delta]$ as the correct confidence interval of α corresponding to the chosen probability level P . If, however, $a^{**} - a^*$ is large, the interval I is unnecessarily large. In such cases we may get a much shorter confidence interval by using some other grouping defined independently of the errors $\epsilon_1, \dots, \epsilon_N; \eta_1, \dots, \eta_N$. For instance if we see that the values x_1, \dots, x_N considered in the order as they have been observed, show a monotonically increasing (or decreasing) tendency, we shall define the group G_1 as the first half, and the group G_2 as the second half of the observations. Though we decide to make this grouping after having observed that the values x_1, \dots, x_N show a clear trend, the grouping can be considered as independent of the errors $\epsilon_1, \dots, \epsilon_N$. In fact, if the range of the error ϵ is small in comparison to the true part X , the trend tendency of the value x_1, \dots, x_N will not be affected by the size of the errors $\epsilon_1, \dots, \epsilon_N$. We may use for the grouping also any other property of the data which is independent of the errors.

The results of the preceding considerations can be summarized as follows:

We use first the grouping according to Rule I, calculate the slope $a' = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2}$ and the corresponding confidence interval $[a' - \delta, a' + \Delta]$ (formula (23)). This confidence interval cannot be considered as exact since the grouping according to Rule I is not completely independent of the errors. In order to take account of this fact, we calculate a^* and a^{**} . If $a^{**} - a^*$ is small, we consider $I = [a^* - \delta, a^{**} + \Delta]$ with practical approximation as the correct confidence interval. If, however, $a^{**} - a^*$ is large, the interval I is unnecessarily large. We can only say that I is a confidence interval corresponding to a probability level greater than or equal to the chosen one. In such cases we should try to use some other grouping defined independently of the errors, which eventually will lead to a considerably shorter confidence interval.

Analogous considerations hold regarding the joint confidence region for α and β . We use the grouping according to Rule I and calculate from (27) the

corresponding confidence region R . If $|a^{**} - a^*|$ and $|b^{**} - b^*|$ are small ($b^* = \bar{y} - a^*\bar{x}$ and $b^{**} = \bar{y} - a^{**}\bar{x}$) we enlarge R to a region \bar{R} corresponding to the fact that a and b may take any values within the intervals $[a^{**}, a^*]$ and $[b^{**}, b^*]$ respectively. The region \bar{R} can be considered with practical approximation as the correct confidence region. If $|a^{**} - a^*|$ or $|b^{**} - b^*|$ is large, we may try some other grouping defined independently of the errors, which may lead to a smaller confidence region. In any case \bar{R} represents a confidence region corresponding to a probability level greater than or equal to the chosen one.

8. Some Remarks on the Consistency of the Estimates of $\alpha, \beta, \sigma_e, \sigma_\eta$. We have shown in section 3 that the given estimates of α, β, σ_e and σ_η are consistent if condition V is satisfied.

If the values x_1, \dots, x_N are not obtained by random sampling, it will in general be possible to define a grouping which is independent of the errors and for which condition V is satisfied. We can sometimes arrange the experiments such that no values of the series x_1, \dots, x_N should be within the interval $[x - c, x + c]$ where x denotes the median of x_1, \dots, x_N and c the range of the error ϵ . In such cases, as we saw, the grouping according to Rule I is independent of the errors. Condition V is certainly satisfied if we group the data according to Rule I.

Let us now consider the case that X_1, \dots, X_N are random variables independently distributed, each having the same distribution. Denote by X a random variable having the same probability distribution as possessed by each of the random variables X_1, \dots, X_N . Assuming that X has a finite second moment, the expression in condition V will approach zero stochastically with $N \rightarrow \infty$ for any grouping defined independently of the values X_1, \dots, X_N . It is possible, however, to define a grouping independent of the errors (but not independent of X_1, \dots, X_N) for which the expression in V does not approach zero, provided that X has the following property: There exists a real value λ such that the probability that X will lie within the interval $[\lambda - c, \lambda + c]$ (c denotes the range of the error ϵ) is zero, the probability that $X > \lambda + c$ is positive, and the probability that $X < \lambda - c$ is positive. The grouping can be defined, for instance, as follows:

The i -th observation (x_i, y_i) belongs to the group G_1 if $x_i \leq \lambda$ and to G_2 if $x_i > \lambda$. We continue the grouping according to this rule up to a value i for which one of the groups G_1, G_2 contains already $N/2$ elements. All further observations belong to the other group.

It is easy to see that the probability is equal to 1 that the relation $x_i \leq \lambda$ is equivalent to the relation $X_i < \lambda - c$ and the relation $x_i > \lambda$ is equivalent to the relation $X_i > \lambda + c$. Hence this grouping is independent of the errors. Since for this grouping condition V is satisfied, our statement is proved.

If X has not the property described above, it may happen that for every grouping defined independently of the errors, the expression in condition V con-

verges always to zero stochastically. Such a case arises for instance if X , ϵ and η are normally distributed.¹⁴ It can be shown that in this case no consistent estimates of the parameters α and β can be given, unless we have some additional information not contained in the data (for instance we know *a priori* the ratio $\sigma_\epsilon/\sigma_\eta$).

9. Structural Relationship and Prediction.¹⁵ The problem discussed in this paper was the question as to how to estimate the relationship between the true parts X and Y . We shall call the relationship between the true parts the structural relationship. The problem of finding the structural relationship must not be confused with the problem of prediction of one variable by means of the other. The problem of prediction can be formulated as follows: We have observed N pairs of values $(x_1, y_1), \dots, (x_N, y_N)$. A new observation on x is given and we have to estimate the corresponding value of y by means of our previous observations $(x_1, y_1), \dots, (x_N, y_N)$. One might think that if we have estimated the structural relationship between X and Y , we may estimate y by the same relationship. That is to say, if the estimated structural relationship is given by $Y = aX + b$, we may estimate y from x by the same formula: $y = ax + b$. This procedure may lead, however, to a biased estimate of y . This is, for instance, the case if X , ϵ and η are normally distributed. It can easily be shown in this case that for any given x the conditional expectation of y is a linear function of x , that the slope of this function is different from the slope of the structural relationship, and that among all unbiased estimates of y which are linear functions of x , the estimate obtained by the method of least squares has the smallest variance. Hence in this case we have to use the least square estimate for purposes of prediction. Even if we would know exactly the structural relationship $Y = \alpha X + \beta$, we would get a biased estimate of y by putting $y = \alpha x + \beta$.

Let us consider now the following example: X is a random variable having a rectangular distribution with the range $[0, 1]$. The random variable ϵ has a rectangular distribution with the range $[-0.1, +0.1]$. For any given x let us denote the conditional expectation of y by $E(y | x)$ and the conditional expectation of X by $E(X | x)$. Then we obviously have

$$E(y | x) = \alpha E(X | x) + \beta.$$

Now let us calculate $E(X | x)$. It is obvious that the joint distribution of X and ϵ is given by the density function:

$$5 \, dX \, d\epsilon,$$

¹⁴ I wish to thank Professor Hotelling for drawing my attention to this case.

¹⁵ I should like to express my thanks to Professor Hotelling for many interesting suggestions and remarks on this subject.

where X can take any value within the interval $[0, 1]$ and ϵ can take any value within $[-0.1, +0.1]$. From this we obtain easily that the joint distribution of x and X is given by the density function

$$5 \, dx \, dX,$$

where x can take any value within the interval $[-0.1, 1.1]$ and X can take any value lying in both intervals $[0, 1]$ and $[x - 0.1, x + 0.1]$ simultaneously. Denote by I_x the common part of these two intervals. Then for any fixed x the relative distribution of X is given by the probability density

$$\frac{dX}{\int_{I_x} dX}.$$

Hence, we have

$$E(X | x) = \frac{\int_{I_x} X \, dX}{\int_{I_x} dX}.$$

We have to consider 3 cases:

$$(1) \quad 0.1 \leq x \leq 0.9.$$

In this case $I_x = [x - 0.1, x + 0.1]$ and

$$E(X | x) = \frac{\int_{x-0.1}^{x+0.1} X \, dX}{\int_{x-0.1}^{x+0.1} dX} = x.$$

$$(2) \quad -0.1 < x \leq 0.1. \quad \text{Then } I_x = [0, x + 0.1] \text{ and}$$

$$E(X | x) = \frac{\int_0^{x+0.1} X \, dX}{\int_0^{x+0.1} dX} = .5x + .05.$$

$$(3) \quad 0.9 \leq x < 1.1. \quad \text{Then } I_x = [x - 0.1, 1] \text{ and}$$

$$E(X | x) = \frac{\int_{x-0.1}' X \, dX}{\int_{x-0.1}' dX} = .5x + .45.$$

Since

$$E(y | x) = \alpha E(X | x) + \beta,$$

we see that the structural relationship gives an unbiased prediction of y from x if $0.1 \leq x \leq 0.9$, but not in the other cases.

The problem of cases for which the structural relationship is appropriate also for purposes of prediction, needs further investigation. I should like to mention a class of cases where the structural relationship has to be used also for prediction. Assume that we have observed N values $(x_1, y_1), \dots, (x_N, y_N)$ of the variables x and y for which the conditions I-IV of section 2 hold. Then we make a new observation on x obtaining the value x' . We assume that the last observation on x has been made under changed conditions such that we are sure that x' does not contain error, i.e. x' is equal to the true part X' . Such a situation may arise for instance if the error ϵ is due to errors of measurement and the last observation has been made with an instrument of great precision for which the error of measurement can be neglected. In such cases the prediction of the corresponding y' has to be made by means of the estimated structural relationship, i.e. we have to put $y' = ax' + b$.

The knowledge of the structural relationship is essential for constructing any theory in the empirical sciences. The laws of the empirical sciences mostly express relationships among a limited number of variables which would prevail exactly if the disturbing influence of a great number of other variables could be eliminated. In our experiments we never succeed in eliminating completely these disturbances. Hence in deducing laws from observations, we have the task of estimating structural relationships.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

A METHOD FOR MINIMIZING THE SUM OF ABSOLUTE VALUES OF DEVIATIONS

BY ROBERT R. SINGLETON

1. Introduction. In the *Philosophical Magazine*, 7th series, May 1930, E. C. Rhodes described a method of computation for the estimation of parameters by minimizing the sum of absolute values of deviations. His is an iterative and recursive method, in the following sense. There is a direct method for minimization with one parameter. Assuming a method for minimization with $n - 1$ parameters, Rhodes imposes a relation between the n parameters (in an n -parameter problem) and finds a restricted minimum by the method for $n - 1$ parameters. In this sense his method is recursive. He then repeats the process, by imposing on the n parameters a new relation determined by the restricted minimum. In this sense his method is iterative. The process is finite, ending when a restricted minimum immediately succeeds itself, indicating a true minimum.

Rhodes' paper presents the method without proof. The purpose of the present paper is to analyze the situation in detail sufficient to indicate proofs for various methods, and to present a new method which reduces the labor of solution by eliminating the recursive feature. The iterative approach is retained. The solution of Rhodes' illustrative problem will be given for comparison between the two methods.

The paper uses geometric terminology and develops to quite an extent the geometry of a surface representing the summed absolute deviations. This seems the clearest means of presenting the relationships. Further analysis of the properties of this surface should lead to an even more direct method for attaining the minimum than the one here presented.

In the writing of the paper, no attention has been given to sets of observations or equations among which a linear dependence may exist. In practice, such a situation almost never occurs. If the need arises, the adjustments which must be made to take care of dependence are in each case fairly obvious.

2. Geometric Analogue of Summed Absolute Deviations. Let n observations on $\nu + 1$ variates be represented by x_α^i, y^i where $i = 1, \dots, n; \alpha = 1, \dots, \nu$. Unless otherwise noted, latin indices have range 1 to n , greek indices, 1 to ν . The summation convention of tensor analysis is used.

¹ The variates are to be statistically related by the linear function¹

$$\hat{y}^i = x_\alpha^i u^\alpha,$$

¹ This includes the linear function with a constant, since a variate $x^i = 1$ may be used.

\hat{y}^i being an estimate of y^i . u^α are to be determined so that $v = \sum_i |\hat{y}^i - y^i|$ is a minimum. Set

$$(1) \quad v^i = x_\alpha^i u^\alpha - y^i$$

and determine functions $e^i(u^\alpha)$ so that $e^i v^i \geq 0$, and $|e^i| = 1$. It is immaterial that e^i is not uniquely determined when u^α satisfies $v^i = 0$. Then $v = \sum_i e^i v^i$ is to be minimized. Using (1),

$$(2) \quad v = x_\alpha u^\alpha - y$$

where

$$x_\alpha = \sum_i e^i x_\alpha^i, \quad y = \sum_i e^i y^i.$$

Consider a Euclidean $(\nu + 1)$ -space, $E_{\nu+1}$, with coordinates u^1, \dots, u^ν, v . The coordinate hyperplane perpendicular to the v -axis will be called E_ν . In $E_{\nu+1}$ each of equations (1) for a particular i represents a ν -plane which intersects E_ν in a $(\nu - 1)$ -plane when $v^i = 0$. Each of the equations

$$(3) \quad v^i = e^i(x_\alpha^i u^\alpha - y^i)$$

represents two half-planes which touch E_ν and each other along the $(\nu - 1)$ -plane given in E_ν by the equation

$$(4) \quad x_\alpha^i u^\alpha - y^i = 0.$$

The functions on the right-hand side of (3) are thus continuous everywhere, and linear in any neighborhood of E_ν , none of whose points satisfies (4). Since a sum of functions continuous and linear in a neighborhood is also continuous and linear in that neighborhood, it follows that the function on the right in (2) is continuous for all u , and linear for every neighborhood of E_ν containing no points which satisfy (4) for any i . Hence

OBSERVATION I: *The surface (S) given in $E_{\nu+1}$ by (2) consists of portions of ν -planes joined together. The projection of these joins on E_ν forms a network of $(\nu - 1)$ -planes determined in E_ν by equations (4).*

3. Existence of a Minimum. Define a "bend of degree r on S " to be the locus of all points on S whose u -coordinates satisfy a set of r independent equations of (4). To each set of r independent equations corresponds a unique bend of degree r .

If a linear relation $u^\alpha = a_\sigma^\alpha \lambda^\sigma + b^\alpha$, $\sigma = 1, \dots, \mu < \nu$, rank $(a_\sigma^\alpha) = \mu$, is imposed on u^α , all the preceding development, reduced in dimension, applies to the new variates $x_\alpha^i a_\sigma^\alpha, y^i - x_\alpha^i b^\alpha$.

OBSERVATION II: *A section of S by a plane of any dimension $d < \nu$ has all the properties of an S -surface of dimension d .*

Since any set of consistent equations selected from (4) determines such a linear relation for u^α , the application of Observation I to any of the bends of S shows that each r -bend consists of linear elements of dimension $\nu - r$, joined

at points which lie on linear elements of lesser dimension. Thus S is a polyhedron. Its faces we term complexes of dimension ν , C_ν , and the linear elements of its edges which lie wholly in bends of degree r , but not of degree $r + 1$ are complexes $C_{\nu-r}$ of dimension $\nu - r$. The boundary of any C_α , $\alpha > 0$, consists of complexes of lesser dimension. The term complex is not restricted to either open or closed complexes.

Since the function $v(u^\alpha)$ of (2) is non-negative, it possesses a greatest lower bound (g.l.b.) g . Since for some number $h > g$, there exists an N such that for all $|u^\alpha| > N$, $v(u^\alpha) > h$, it follows that for some closed neighborhood of E , the g.l.b. of v is g . Since v is continuous everywhere it attains its g.l.b., and so S has minimum points. Since the minimum of any complex not parallel to E_r , lies on its boundary, and the boundary consists of complexes, it follows that the minimum points of S consist of C_0 's and/or entire complexes of dimension > 0 which are parallel to E_r . The next section will show that S has a unique minimum complex (including of course its boundary complexes) and furthermore is cup-shaped.

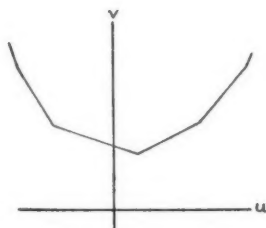


FIG. 1

4. Convexity Property; Uniqueness of the Minimum. Consider $\nu = 1$ in the preceding treatment (and for convenience not written). S looks generally like Fig. 1. The slope changes only where an equation of (4) has a root. Suppose the point is u_0 , and $x^1 u_0 - y^1 = 0$. From (3), since $v^1 \geq 0$, it follows that $e^1 x^1 < 0$ for $u < u_0$, $e^1 x^1 > 0$ for $u > u_0$. Since in (2) $x = \sum e^i x^i$, and since for h sufficiently small and $u_0 - h < u < u_0 + h$ the only e to change value² is e^1 , we have that

$$x(u_1) + 2|e^1 x^1| = x(u_2)$$

where

$$u_0 - h < u_1 < u_0 < u_2 < u_0 + h.$$

Hence the slope is a monotonic increasing step function. Since for u sufficiently small all $e^i x^i < 0$, and for u sufficiently large all $e^i x^i > 0$, at some intermediate point or points either the slope is zero or it changes from negative to

² The e 's corresponding to equations proportional to equation (1) also change value at x_0 . This does not destroy the argument.

positive without becoming zero. In the first case a single closed C_1 is the minimum complex; in the second, a C_0 . In either case the curve given by (2) when $\nu = 1$ is concave upward and has just one minimum complex, except for complexes of lesser dimension constituting the boundary of this complex. An obvious consequence is

LEMMA I. *The set of points u for which v is less than some number N form a convex point set.*

This result is easily extended to the general dimension ν . If for any two points u_1, u_2 of E_r , $v(u_1) < N$ and $v(u_2) < N$, the plane in E_{r+1} given by $u^\alpha = u_1^\alpha + \lambda(u_2^\alpha - u_1^\alpha)$ makes a one-dimensional section of S . By Observation II, the points u lying on the projection of this section on E_r have the property of Lemma I and of course lie on the straight line joining u_1 and u_2 . This is the property required for a convex point set. Hence

THEOREM I. *The set of points u^α of E_r for which $v(u^\alpha)$ as given by (2) is less than a fixed quantity form a convex point set.*

From this it follows immediately that there is a unique minimum complex. It is appropriate here to point out that no two complexes can be contained in a single plane of the same dimension. This follows from the equation giving monotonicity of slope in one dimension, and Observation II.

5. Gradient Directions. From here on the treatment will be of v as a function defined on E_r , and the equations will represent objects in E_r , unless otherwise stated. Complex and Bend also will refer to the projections on E_r of the complexes and bends of S . For a single-valued function defined on E_r the gradient at a point is the projection of a normal to the surface representing the function in E_{r+1} . If the function is defined only over a subspace of E_r possessing derivatives, the gradient will be required also to be tangent to the subspace. This is sufficient to determine a unique direction, and preserves the property that for an infinitesimal displacement in any direction the value of the function decreases most rapidly in the direction of the gradient. Here gradient is taken negative to its usual sense.

A point u lying on a C_r but not on a C_{r-1} will have a gradient in C_r and also in each higher-dimensional complex on whose boundary C_r lies. If the gradient for u as a point of C_{r+k} points into C_{r+k} (remembering that u lies on the boundary) this will be called a usable gradient. In the case of the greatest k for which there exists a usable gradient, there exists but one C_{r+k} providing such a gradient, and that gradient is the "best" gradient; that is, of all directions in E_r it provides the direction of most rapid decrease of the function v . This follows from Theorem I. Furthermore, all complexes of lesser dimension providing usable gradients lie on the boundary of this C_{r+k} . In fact

THEOREM II. *If for a point u on C_r , two complexes C_s and C'_s , $s > r$, lying in different bends of degree $\nu - s$ but incident at C_r , both provide usable gradients for u , then the complex C_{s+1} on whose boundary lie both C_s and C'_s also provides a usable gradient for u .*

This follows from Theorem I. Select u_1 on the gradient in C_s , u_2 on the gradient in C'_s , for which $v(u_1) = v(u_2)$. The join of u_1 and u_2 lies in C_{s+1} , and for some point, u_3 on this join, $v(u_3)$ is less than $v(u_1) = v(u_2)$. Also, the distance $\overline{u u_3}$ is less than at least one of $\overline{u u_1}$, $\overline{u u_2}$. Hence C_{s+1} must contain a usable gradient.

6. Selection of Best Gradient at Bends. The direction of the gradient for a point u_0 considered as lying on a C , is given by

$$(5) \quad g^a = -x_a(u_0) = -\sum_i e^i(u_0)x_a^i.$$

If u_0 lies in the interior of a face, this is unique. If u_0 lies in a bend, so that some e^i are not determined, the g^a for each face is found by selecting the indeterminate e 's as $+1$ or -1 , according to the face being considered.

For a point u_0 considered as lying on a bend of degree r , given by r independent equations of (4):

$$(6) \quad x_a^\lambda u^\alpha - y^\lambda = 0, \quad (\lambda = 1, \dots, r),$$

the gradient for a particular C_{s-r} , determined by the conditions at the beginning of section 5, is

$$(7) \quad g^a = x_a^\lambda k_\lambda - x_a$$

where k_λ satisfies

$$\sum_a x_a^\mu x_a^\lambda k_\lambda = \sum_a x_a^\mu x_a, \quad (\mu = 1, \dots, r)$$

and x_a is as given in (2), the choice of sign for the indeterminate e^λ ($\lambda = 1, \dots, r$) being immaterial. They may, in fact, be taken as 0 in this instance.

For a point u_0^α lying on an r -bend given by (6), to determine which complex contains the best gradient, each $(r-1)$ -bend incident on the r -bend at u_0 is tested for a usable gradient. Theorem II then determines the complex containing the best gradient.

There are $2r$ such complexes incident at u_0 , given by the r sets of equations selected from (6):

$$(8) \quad (\lambda): x_a^\sigma u^\alpha - y^\sigma = 0 \quad \begin{aligned} &(\sigma = 1, \dots, \lambda-1, \lambda+1, \dots, r) \\ &(\lambda = 1, \dots, r). \end{aligned}$$

The two complexes lying in the same $(r-1)$ -bend have the same equations in (8), but are distinguished later by $e^\lambda(u_0)$ for the omitted equation being taken first $+1$, then -1 .

The gradient for the λ th pair of complexes is

$$g_\lambda^a = x_a^\sigma k_\sigma - x_a$$

similar to (7), but not identical. For $e^\lambda = +1$ in determining x_a , we have $g_{\lambda+}^a$, and for $e^\lambda = -1$, $g_{\lambda-}^a$. We restrict the consideration to $e^\lambda = +1$.

The line in the direction of greatest slope is then

$$u^\alpha = u_0^\alpha + g_{\lambda+}^\alpha t.$$

Now u_0 is here considered lying on the complex given by (8 λ) with $e^\lambda = +1$. In order that $g_{\lambda+}^\alpha$ point into this face, the deviation for the λ th observation must exceed 0 when $t > 0$; otherwise, for a displacement in the direction of $g_{\lambda+}^\alpha$, e^λ changes sign immediately and the course is in the other complex. This deviation is

$$v^\lambda = x_\alpha^\lambda u^\alpha - y^\lambda = x_\alpha^\lambda u_0^\alpha - y^\lambda + x_\alpha^\lambda g_{\lambda+}^\alpha t = x_\alpha^\lambda g_{\lambda+}^\alpha t.$$

Had $g_{\lambda-}^\alpha$ been used, this deviation must be less than 0. Hence a necessary and sufficient condition that a complex given by (8) with either choice of e^λ possess a usable gradient is

$$(9) \quad \Phi_\lambda = e^\lambda [\sum_\alpha x_\alpha^\lambda x_\alpha^\sigma k_\sigma - \sum_\alpha x_\alpha^\lambda x_\alpha] > 0.$$

For $r = 1$ the condition is given by (9) with the first sum merely omitted. $\Phi_{\lambda+}$ and $\Phi_{\lambda-}$ cannot both exceed 0.

When all sets of equations (8 λ) are tested by (9) the equations common to all sets possessing a usable gradient determine the complex with the best gradient, retaining the values of e for which (9) was satisfied.

7. Property of the Minimum Point. For a minimum point, given by (6) with $r = \nu$, all Φ_λ must be negative. Define $X^{\beta\gamma} = \sum_\alpha x_\alpha^\beta x_\alpha^\gamma$ and $X^{\beta 0} = \sum_\alpha x_\alpha^\beta x_\alpha$ for convenience. Then in (9), the numbers k_σ , -1 are seen from their definition in (7) to be proportional to the cofactors of the λ th row of the matrix $(X^{\mu\sigma}, X^{\mu 0})$, μ having the same range as λ . Thus $\Phi_{\lambda+} = c \text{Det}(X^{\mu\sigma}, X_+^{\mu 0})$, and $\Phi_{\lambda-} = -c \text{Det}(X^{\mu\sigma}, X_-^{\mu 0})$, where in the first case $X^{\mu 0}$ is determined with $e^\lambda = +1$, in the second with $e^\lambda = -1$. The factor of proportionality, c , must be the same since $X^{\mu\sigma}$ is unaffected by change of e^λ . Now let $X^\mu = \sum_\alpha x_\alpha^\mu x_\alpha^*$ where $x_\alpha^* = \sum_k e^k x_\alpha^k$, the range of k omitting the range of λ . Then

$$\Phi_{\lambda+} = c [\text{Det}(X^{\mu\sigma}, X^\mu) + \text{Det}(X^{\mu\sigma}, X^{\mu\lambda})]$$

and

$$\Phi_{\lambda-} = -c [\text{Det}(X^{\mu\sigma}, X^\mu) - \text{Det}(X^{\mu\sigma}, X^{\mu\lambda})].$$

Hence

$$\Phi_{\lambda+}\Phi_{\lambda-} = -c^2 \{[\text{Det}(X^{\mu\sigma}, X^\mu)]^2 - [\text{Det}(X^{\mu\sigma}, X^{\mu\lambda})]^2\}.$$

Now let A represent the square matrix (x_α^λ) , α giving the rows and λ the columns. Let B_λ represent the matrix formed from A by replacing the λ th column by x_α^* . Then

$$\begin{aligned} \Phi_{\lambda+}\Phi_{\lambda-} &= -c^2 [\text{Det}^2(A'B_\lambda) - \text{Det}^2(A'A)] \\ &= -c^2 \text{Det}^2 A (\text{Det}^2 B_\lambda - \text{Det}^2 A) \end{aligned}$$

and this will have the same sign as

$$\Psi_{\lambda} = |\text{Det}(A)| - |\text{Det}(B_{\lambda})|.$$

Since $\Phi_{\lambda+}$ and $\Phi_{\lambda-}$ are never both positive, and at the minimum are both negative for all λ , at the minimum all $\Psi_{\lambda} > 0$. To determine all Ψ_{λ} together, let, in matrix notation, $z' = (z_1, \dots, z_r)$ and $x^{*'} = (x_1^*, \dots, x_r^*)$ where x_a^* were defined previously. Determine z as the solution of $Az = x^*$. Then $|\text{Det}(B_{\lambda})|$ are equal to $|z_{\lambda}| |\text{Det}(A)|$. Hence a necessary and sufficient condition that $\Psi_{\lambda} > 0$ for all λ is that all $|z_{\lambda}|$ be less than one. Hence

THEOREM III: *If a zero-complex is given by a set of equations whose matrix is M , a necessary and sufficient condition that the complex be a unique minimum is that the solutions of $M'z = x^*$ be all less than one in absolute value. If k of the solutions are equal to one in absolute value, and the rest are less than one, the minimum is a complex of dimension k with the zero-complex as one of its corners.*

The last statement follows since if one solution is 1 in absolute value, a corresponding $\Phi_{\lambda} = 0$, and hence no gradient, usable or not, exists. Thus the corresponding complex is parallel to E_r .

8. Minimization for One Dimension. A method for minimization of (2) when there is just one parameter evolves from the monotonicity of slope in that case. Suppose the variates are w^i and z^i , and (1) is

$$(10) \quad v^i = w^i t - z^i.$$

Suppose the variates are arranged in order of z^i/w^i , starting with the smallest. The slope of the r th segment (Fig. 1) from the left is

$$\sum_{i=1}^r |w^i| - \sum_{i=r+1}^n |w^i|.$$

The minimum occurs when the slope is 0 or changes from negative to positive; that is, when the first sum equals or exceeds the second; or when the first sum equals or exceeds half the total. This is a standard computation. If the change takes place when $r = k$, then $t = z^k/w^k$ is the value of t giving the minimum.

9. Minimization Procedure for $\nu + 1$ Dimensions. For any continuous function with unique minimum and having the property of Theorem I, the following holds. Let u_0 be any point of E_r . Let $u_{i+1} = u_i + \lambda_i t_i$, where λ_i is any direction chosen at random and t_i is the value of t for which the function attains a minimum on the curve $u = u_i + \lambda_i t$. Then the probability is one that $\lim_{i \rightarrow \infty} u_i = u_1$, where u_1 is a minimum point for the function. If λ_i is taken always as the gradient of u_i , such a procedure is called the "method of steepest descent" for approaching the minimum point.

Usually the limit is never attained. In this case, however, the minimum is

TABLE I
Method of Steepest Descent Applied

Data					First Restricted Minimum					Second Restricted Minimum					Third (absolute) Minimum					for test
(2)	(3)	(4)	(5)		(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
x_1^i	x_2^i	x_3^i	y^i		x_0^i	$e^i(u_0)$	w_0^i	$\frac{z_0^i}{w_0^i}$	rank of (9)	z_1^i	e_1^i	w_1^i	$\frac{z_1^i}{w_1^i}$	rank of (14)	z_2^i	e_2^i	w_2^i	$\frac{z_2^i}{w_2^i}$	rank of (10)	i
1	-8	64	28		-78	-	-5165	.015	16	-47.044	-	-1149.570	-.0409	6	-92.773	-	-285.180	.3253	15	1
2	-7	49	-39		14	+	-3917	-.004	8	37.476	+	-942.102	.0398	12	0		0			2
3	-6	36	-40		36	+	-2841	-.013	6	53.027	+	-753.324	.0704	14	23.060	+	247.156	.0933	13	3
4	-5	25	-1		14	+	-1937	-.007	7	25.609	+	-583.236	.0439	13	2.408	+	456.288	.0053	8	4
5	-4	16	23		3	+	-1205	-.002	9	10.222	+	-431.838	.9237	11	-6.956	-	627.396	-.0111	2	5
6	-3	9	34		1	+	-645	-.0015	10	4.866	+	-299.130	.0163	10	-7.033	-	760.480	-.0092	3	6
7	-2	4	25		15	+	-257	-.058	2	16.540	+	-185.112	.0894	15	9.177	+	855.540	.0107	9	7
8	-1	1	40		1	+	-41	-.024	3	1.246	+	-89.784	.0139	8	-2.326	-	912.576	-.0025	5	8
9	0	0	43		-5	-	3	-1.67	1	-5.018	-	-13.146	-.3817	2	-5.541	-	931.588	-.0059	4	9
10	1	1	29		2	+	-125	-.016	4	2.749	+	44.802	-.0614	5	4.531	+	912.576	.0050	7	10
11	2	4	14		6	+	-425	-.014	5	8.547	+	84.060	-.1017	4	11.891	+	855.540	.0139	10	11
12	3	9	12		-7	-	-897	.008	14	-1.624	-	104.628	.0155	9	2.538	+	760.480	.0033	6	12
13	4	16	-16		2	+	-1541	-.001	11	11.236	+	106.506	-.1055	3	15.472	+	627.396	.0247	12	13
14	5	25	-14		-23	-	-2357	.010	15	-8.874	-	89.694	.0989	16	-5.306	-	456.288	-.0116	1	14
15	6	36	-46		-18	-	-3345	.005	12	2.048	+	54.192	-.0378	7	4.203	+	247.156	.0170	11	15
16	7	49	-68		-27	-	-4505	.006	13	0		0			0		0			16
17	8	64	-24		-106	-	-5837	.018	17	-71.017	-	-72.882	-.9744	1	-73.916	-	-285.180	.2592	14	17
Σ	17	0	408	0	-170	-	-35037			39.990	-	-4036.242			-120.570		8080.100			Σ

attained. The minimum can be approached as closely as desired, hence a complex incident on the minimum is reached. But the convex point sets of Theorem I surrounding the minimum complex are all similar convex polyhedrons in E_r , whose corresponding faces are parallel, and the gradients at points on a bend cannot point into a higher dimensional complex on the bend. Hence the sequence of points lie on bends of successively greater degree, and must eventually attain the minimum complex.

TABLE II

Points u_k

$u_{k+1}^a = u_k^a + g_k^a t_k$		
u_0	$= (38, -5, -2)$	
u_1	$= (37.98202, -4.74828, -1.48457)$	
u_2	$= (37.45908, -2.07142, -1.85631)$	
u_3	$= (32.83333, -2.07142, -1.76191)$	

TABLE III

Computation of $t_k = z_k/w_k$

$\Sigma w_k $	in order of col.	exceeds	at $i =$	hence $t_k =$
$\Sigma w_0 $	(10)	17521	16	.00599334
$\Sigma w_1 $	(15)	2502	2	.0397792
$\Sigma w_2 $	(20)	4610	10	.00496545

TABLE IV

Gradients g_k^a for column $(5k + 8)$

k	g_k^1	g_k^2	g_k^3
0	-3	42	86
1	-13146	67293	-9345
2	-931588	0	19012

The computational procedure is as follows:

1. Select a point u_0 .
 2. Determine the gradient g_0^a from (5).
 3. Compute $w_0^i = x_a^i g_0^a$, $z_0^i = y^i - x_a^i u_0^a$.
 4. Determine t_0 by the method of section 8.
 5. Compute $u_1^a = u_0^a + g_0^a t_0$.
 6. Determine the complex containing the best gradient by (9), and the gradient g_1^a by (7).
- and so proceed to the minimum. This may be finally tested by Theorem III.

Step 5 is unnecessary, since the only use for u_1^a is to determine $e^i(u_1)$. But $e^i(u_1) = e^i(t_0)$, the latter referring to the computation in step 4. Also, after the first step, it is easier to compute z^i by

$$z_{k+1}^i = z_k^i - w_k^i t_k.$$

10. **Example.** The computation for (9) is not so great as it would seem, since some of the work is duplication and some must be computed anyway for the gradient. Even so, for $r \geq 3$ it becomes, perhaps, more arduous than its contribution would seem to justify. For $\nu \geq 4$ it is recommended that the test of (9) be omitted for points on bends of third degree or greater, and the final test of Theorem III be applied at the end of the work. If this test shows the minimum has not been reached, the complex in which lies the best gradient will be indicated at the same time.

The minimum number of steps is 0. The maximum number is tremendous but finite. The expected number is probably a little greater than ν .

In Tables I to IV, the method is applied to the problem used by Rhodes to illustrate his method. The independent variates are shown in columns (2), (3), (4), Table I, the dependent variate in column (5). The only other original datum is the initial point, selected by guess, shown in line 1, Table II. Since slightly different formulas were used in the computation, the signs of cols. (6), (8), (11), (16), (18) are reversed, and the gradients in Table IV are multiplied by constants. As they are used only for directions, this does not matter.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

A STUDY OF A UNIVERSE OF n FINITE POPULATIONS WITH APPLICATION TO MOMENT-FUNCTION ADJUSTMENTS FOR GROUPED DATA

BY JOSEPH A. PIERCE

The object of this paper is to study the case of a universe of n finite populations, considering both the expectations of population moment-functions and the moments of sample moments, and to make applications of the results which may be of interest to mathematical statisticians. The sampling formulas which are derived reduce to the usual infinite or finite sampling formulas, under appropriate assumptions. Also a method is given whereby finite sampling formulas may be transformed into the corresponding infinite sampling formulas.

The general methods and formulas which are given in Part I for the expectations of population moment-functions are used, in Part II, to find the expectations of moments of a distribution of discrete data grouped in " k groupings of k ".

I. A STUDY OF A UNIVERSE OF n FINITE POPULATIONS

Let ${}_nU_N$ be a universe composed of the set of populations ${}_rX$, ($r = 1, 2, \dots, n$) each population ${}_rX$ consisting of a finite number of discrete variates ${}_rx_i$, ($i = 1, 2, \dots, N$), ($N > n$). The t th moment of ${}_rX$ is denoted by ${}_r\mu_t$. The t th central moment of ${}_rX$ is denoted by ${}_r\bar{\mu}_t$. The t th moment and the t th central moment of ${}_nU_N$ are respectively denoted by μ_t and $\bar{\mu}_t$. The expected value of a variable y is denoted by $E(y)$. We have

$$\begin{aligned} {}_r\mu_t &= E({}_rx_i^t) = \frac{1}{N} \sum_{i=1}^N {}_rx_i^t, & {}_r\bar{\mu}_t &= E({}_rx_i - {}_r\mu_1)^t = \frac{1}{N} \sum_{i=1}^N ({}_rx_i - {}_r\mu_1)^t, \\ (1.1) \quad \mu_{1:\mu_t} &= E({}_r\mu_t) = \frac{1}{n} \sum_{r=1}^n {}_r\mu_t, & \mu_{1:\bar{\mu}_t} &= E({}_r\bar{\mu}_t) = \frac{1}{n} \sum_{r=1}^n {}_r\bar{\mu}_t, \end{aligned}$$

$$\begin{aligned} \mu_{s_1 s_2 \dots s_v; \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}} &= E({}_r\mu_{t_1}^{s_1} {}_r\mu_{t_2}^{s_2} \dots {}_r\mu_{t_v}^{s_v}), \\ \mu_{s_1 s_2 \dots s_v; \bar{\mu}_{t_1} \bar{\mu}_{t_2} \dots \bar{\mu}_{t_v}} &= E({}_r\bar{\mu}_{t_1}^{s_1} {}_r\bar{\mu}_{t_2}^{s_2} \dots {}_r\bar{\mu}_{t_v}^{s_v}). \end{aligned}$$

We also note that $\mu_{s_1 s_2 \dots s_v; \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}}$ may be written $\mu_{111 \dots 1; t \mu_{t_1}^{s_1} \mu_{t_2}^{s_2} \dots \mu_{t_v}^{s_v}}$.

1. The expected value of moments and central moments. It follows easily from (1.1) that

$$(1.2) \quad \mu_{1:\mu_t} = \mu_t.$$

From the usual formula for central moments in terms of moments, we get

$$(1.3) \quad \mu_{1;\bar{\mu}_t} = \sum_{i=0}^t (-1)^i \binom{t}{i} \mu_{i1;\mu_1\mu_{t-1}}.$$

Terms of the form $\mu_{i1;\mu_1\mu_{t-1}}$ may be evaluated by use of the well known formulas [20; p. 58] for changing from moments to central moments in the case of a multivariate distribution. Two of these formulas are given below.

$$(1.4) \quad \begin{aligned} \bar{\mu}_{11;\mu_a\mu_b} &= \mu_{11;\mu_a\mu_b} - \mu_{10;\mu_a\mu_b}\mu_{01;\mu_a\mu_b} \\ \bar{\mu}_{111;\mu_a\mu_b\mu_c} &= \mu_{111;\mu_a\mu_b\mu_c} - \mu_{110;\mu_a\mu_b\mu_c}\mu_{001;\mu_a\mu_b\mu_c} \\ &\quad - \mu_{101;\mu_a\mu_b\mu_c}\mu_{010;\mu_a\mu_b\mu_c} - \mu_{011;\mu_a\mu_b\mu_c}\mu_{100;\mu_a\mu_b\mu_c} \\ &\quad + 2\mu_{100;\mu_a\mu_b\mu_c}\mu_{010;\mu_a\mu_b\mu_c}\mu_{001;\mu_a\mu_b\mu_c}. \end{aligned}$$

We find that

$$(1.5) \quad \mu_{i1;\mu_1\mu_{t-1}} = \sum \frac{i!}{p_1!p_2!r_1!r_2!} \bar{\mu}_{p_1r_1;\mu_1\mu_{t-1}} \mu_{1;\mu_1}^{p_2} \mu_{1;\mu_{t-1}}^{r_2},$$

where p_1p_2 is a two-part partition of i and $r_1 + r_2 = 1$.

Using (1.3) and (1.5), we get

$$(1.6) \quad \mu_{1;\bar{\mu}_2} = \bar{\mu}_2 - \bar{\mu}_{2;\mu_1}.$$

$$(1.7) \quad \mu_{1;\bar{\mu}_3} = \bar{\mu}_3 - 3\bar{\mu}_{11;\mu_1\mu_2} + 6\mu_1\bar{\mu}_{2;\mu_1} + 2\bar{\mu}_{3;\mu_1}.$$

$$(1.8) \quad \begin{aligned} \mu_{1;\bar{\mu}_4} &= \bar{\mu}_4 + 6(\bar{\mu}_2 - 2\mu_1^2) \bar{\mu}_{2;\mu_1} - 12\mu_1\bar{\mu}_{3;\mu_1} + 12\mu_1\bar{\mu}_{11;\mu_1\mu_2} \\ &\quad - 4\bar{\mu}_{11;\mu_1\mu_3} + 6\bar{\mu}_{21;\mu_1\mu_2} - 3\bar{\mu}_{4;\mu_1}. \end{aligned}$$

etc.

If the n populations are identical, it is evident from the definition of $\mu_{1;\bar{\mu}_t}$ that, for all finite t ,

$$\mu_{1;\bar{\mu}_t} = \bar{\mu}_t.$$

2. The expected value of Thiele seminvariants. If the t th Thiele seminvariant is denoted by λ_t , then

$$(1.9) \quad \mu_{1;\lambda_t} = \sum \frac{(-1)^{p-1} t! (\rho - 1)!}{s_1! s_2! \dots s_v! (2!)^{s_2} (3!)^{s_3} \dots (v!)^{s_v}} \mu_{s_1 s_2 \dots s_v; \mu_1 \mu_2 \dots \mu_v},$$

the summation being taken for all positive integers s_i ($i = 1, 2, \dots, v$), for which

$$\rho = \sum_{i=1}^v s_i, \quad t = \sum_{i=1}^v i s_i.$$

Terms of the form $\mu_{s_1 s_2 \dots s_v; \mu_1 \mu_2 \dots \mu_v}$ are evaluated by (1.4). We have

$$(1.10) \quad \mu_{1;\lambda_2} = \lambda_2 - \bar{\mu}_{2;\mu_1}.$$

$$(1.11) \quad \mu_{1:\lambda_3} = \lambda_3 - 3\bar{\mu}_{11:\mu_1\mu_2} + 6\lambda_1\bar{\mu}_{2:\mu_1} + 2\bar{\mu}_{3:\mu_1}.$$

$$(1.12) \quad \begin{aligned} \mu_{1:\lambda_4} = & \lambda_4 + 12[\lambda_2 - 2\lambda_1^2] \bar{\mu}_{2:\mu_1} - 24\lambda_1\bar{\mu}_{3:\mu_1} + 24\lambda_1\bar{\mu}_{11:\mu_1\mu_2} \\ & - 4\bar{\mu}_{11:\mu_1\mu_3} + 12\bar{\mu}_{21:\mu_1\mu_2} - 3\bar{\mu}_{2:\mu_2} - 6\bar{\mu}_{4:\mu_1}. \end{aligned}$$

etc.

If the n populations are identical then, for all finite t ,

$$\mu_{1:\lambda_t} = \lambda_t.$$

3. Generalized sampling. It follows from definition that all rational isobaric moment-functions have the property that they may be expressed in terms of power sums and power product sums with certain coefficients. Of the power sums and power product sums which enter a sampling formula only the power product sums take different forms depending on the law of variate selection. Now, there are two possible courses which may be followed by one who wishes to derive sampling formulas for the case of a single population.

1. One may decide in advance on the law which he wishes to govern the selection of variates which enter the sample. Then he may apply this law in the evaluation, in terms of moments, of every power product term as it occurs in each formula which is derived.

2. One may derive the formulas for sampling under the condition that the law is unspecified, thereby obtaining formulas which are capable of being interpreted in terms of laws that are decided upon later.

We illustrate the two possible courses by considering the formula,

$$(1.13) \quad \bar{\mu}_{2:z} = \frac{r}{s} \Sigma \bar{x}^2 + \frac{2r(r-1)}{s(s-1)} \Sigma \bar{x}_i \bar{x}_j,$$

which Carver [12; p. 102] obtains for the case of finite sampling without replacements. Here r = the number in the sample, s = the number in the parent population and z_i = the algebraic sum of the variates of i th sample. Later, by evaluating $\Sigma \bar{x}^2$ and $\Sigma \bar{x}_i \bar{x}_j$ in terms of moments, he finds

$$(1.14) \quad \bar{\mu}_{2:z} = \frac{r(s-r)}{s-1} \bar{\mu}_{2:z}.$$

(It should be noted that Carver [12; p. 115] obtained the corresponding formula for infinite sampling by letting $s \rightarrow \infty$).

The preceding development is entirely in accord with the first of the courses stated above. It is also the standard procedure and is the course followed by such writers as Isserles [2], Neyman [6], Church [7], Pepper [11] and Dwyer [20], in deriving finite sampling formulas. Also, it is the course followed by such authors as "Student" [1], Tchouproff [3], Church [5], Craig [9], Fisher [10], and Georgesque [13] for the case of sampling from an infinite population.

However, in (1.13), it is possible to employ the definition,

$$\frac{2}{s(s-1)} \sum \bar{x}_i \bar{x}_j = \bar{\mu}_{1,1}.$$

Then (1.14) becomes

$$(1.15) \quad \bar{\mu}_{2:z} = r\bar{\mu}_2 + r(r-1)\bar{\mu}_{1,1}.$$

Formula (1.15) may be interpreted as holding for either finite or infinite sampling, depending on the interpretation which is given to $\bar{\mu}_{1,1}$. It may be easily shown that, if the sampling is from a limited supply, $\bar{\mu}_{1,1} = \frac{-1}{s-1} \bar{\mu}_2$ and (1.15) reduces to (1.14). If the sampling is from an infinite supply, $\bar{\mu}_{1,1}$ becomes $\bar{\mu}_1^2$ and therefore

$$\bar{\mu}_{2:z} = r\bar{\mu}_{2:z},$$

which is the formula [12; p. 115] that corresponds, in the infinite case, to (1.14).

Thus, either of the two courses is possible in the case of sampling from a single population. However, if one wishes to get general formulas which hold for both infinite and finite sampling, he should follow the second course. Similarly, in order to obtain generalized sampling formulas where the relations between the variates are unspecified and the populations are assumed to be different, the second course should be followed.

It appears that Tchouproff [3], [4] was the first to approach the sampling problem from such a general point of view. However, his methods of derivation are quite complicated and his results, in general, are difficult to apply to a given problem [5], [8].

Samples of n are formed from ${}_nU_N$ by choosing one variate from each of the n populations. A typical sample is

$$1x_{i_1}, 2x_{i_2}, 3x_{i_3}, \dots, rx_{i_r}, \dots, nx_{i_n}.$$

We define [4; p. 472]

$$(1.16) \quad \frac{1}{k} \sum_{\substack{i_1, i_2, \dots, i_v=1 \\ r_j \neq r_k}}^n r_1 x_{i_1}^{t_1} r_2 x_{i_2}^{t_2} \dots r_v x_{i_v}^{t_v} = E(r_1 x_{i_1}^{t_1} r_2 x_{i_2}^{t_2} \dots r_v x_{i_v}^{t_v}) \\ = r_1 r_2 \dots r_v \mu_{t_1 t_2 \dots t_v}, \\ \frac{v}{n^{(v)}} \sum_{r_i \neq r_j} r_1 r_2 \dots r_v \mu_{t_1 t_2 \dots t_v} = \frac{1}{n^{(v)}} S_v r_1 r_2 \dots r_v \mu_{t_1 t_2 \dots t_v} = \mu_{t_1 t_2 \dots t_v},$$

where k represents the number of possible terms of the given form; S_v means v times the sum for unequal values of r_1, r_2, \dots, r_v and $n^{(v)} = n(n-1) \dots (n-v+1)$.

4. Moments and product moments of sample moments. The t th moment of the j th sample is denoted by ${}_j m_t$. The s th moment of ${}_j m_t$ for all j is denoted by $'\mu_{s; m_t}$ where the prime indicates that the moments of the universe are measured about a fixed point. It follows that

$$(1.17) \quad {}_j m_t = \frac{1}{n} \sum_{r=1}^n r x_{i_r}^t \quad \text{and} \quad {}' \mu_{s:m_t} = E[{}_j m_t]^s.$$

Also, the general product moment, in which the variates of both the sample and the universe are measured about a fixed point, is defined by

$$(1.18) \quad {}' \mu_{s_1 s_2 \dots s_v; m_{t_1} m_{t_2} \dots m_{t_v}} = E[{}_j m_{t_1}^{s_1} {}_j m_{t_2}^{s_2} \dots {}_j m_{t_v}^{s_v}].$$

As an illustration of the methods used to derive the formulas of this section, consider a special case of (1.18) when $s_1 = 2$ and $s_i = 0$, ($i = 2, 3, \dots, v$). Then

$$\begin{aligned} {}' \mu_{2:m_t} &= \frac{1}{n^2} E \left[\sum_{r=1}^n r x_{i_r}^t \right]^2 \\ &= \frac{1}{n^2} E \left[\sum_{r=1}^n r x_{i_r}^{2t} + S_2 r_1 x_{i_{r_1}}^t r_2 x_{i_{r_2}}^t \right] \\ &= \frac{1}{n^2} \left[\sum_{r=1}^n r \mu_{2t} + S_2 r_1 r_2 \mu_{t,t} \right]. \end{aligned}$$

Therefore, by (1.1), (1.2) and (1.16), we get

$$(1.19) \quad {}' \mu_{2:m_t} = \frac{1}{n^2} [n \mu_{2t} + n^{(2)} \mu_{t,t}].$$

Using the formulas [20; p. 34] relating products of power sums and power products to expand expressions of the type $E({}_j m_{t_1}^{s_1} {}_j m_{t_2}^{s_2} \dots {}_j m_{t_v}^{s_v})$, we give, in the tables below, formulas for moments and product moments of sample moments through weight six. The number in a cell and the coefficient, in the same column, at the top of the table should be taken as the coefficient of the moment which is found in the same vertical division. The coefficients in the vertical division are coefficients of the entire right members of the formulas for the respective moments.

Terms of the form $\mu_{t_1 t_2 \dots t_r}$, if $t_1 = t_2 = \dots = t_r = t$, are sometimes written $\mu_{t^r, t_{r+1} \dots t_v}$.

The numbers in the cells of the tables are identical with the numbers in the cells of the tables given by Dwyer [19; p. 30] for the expected value of partition products.

5. Moments of central moments of samples of n . The t th central moment of the j th sample is denoted by ${}_j \bar{m}_t$. Then,

$$(1.20) \quad {}_j \bar{m}_t = \frac{1}{n} \sum_{r=1}^n (r x_{i_r} - {}_j m_1)^t$$

and

$$(1.21) \quad {}' \mu_{s:\bar{m}_t} = E \left[\frac{1}{n} \sum_{r=1}^n (r x_{i_r} - {}_j m_1)^t \right]^s.$$

TABLE I

(1)			(2)				(3)				
	Coef.	n		Coef.	n	$n^{(2)}$		Coef.	n	$n^{(2)}$	$n^{(3)}$
		μ_1			μ_2	μ_1^2			μ_3	$\mu_{2,1}$	μ_1^3
$\mu_{1:m_1}$	n^{-1}	1		$\mu_{1:m_2}$	n^{-1}	1		$\mu_{1:m_3}$	n^{-1}	1	
				$\mu_{2:m_1}$	n^{-2}	1	1	$\mu_{11:m_1m_2}$	n^{-2}	1	1
								$\mu_{3:m_1}$	n^{-3}	1	3
											1

(5)								
	Coef.	n	$n^{(2)}$	$n^{(2)}$	$n^{(3)}$	$n^{(3)}$	$n^{(4)}$	$n^{(5)}$
		μ_5	$\mu_{4,1}$	$\mu_{3,2}$	$\mu_{3,1^2}$	$\mu_{2^2,1}$	$\mu_{2,1^2}$	μ_{1^5}
$\mu_{1:m_5}$	n^{-1}	1						
$\mu_{11:m_1m_4}$	n^{-2}	1	1					
$\mu_{11:m_2m_3}$	n^{-2}	1		1				
$\mu_{21:m_1m_3}$	n^{-3}	1	2	1	1			
$\mu_{12:m_1m_2}$	n^{-3}	1	1	2		1		
$\mu_{31:m_1m_2}$	n^{-4}	1	3	4	3	3	1	
$\mu_{5:m_1}$	n^{-5}	1	5	10	10	15	10	1

(4)						
	Coef.	n	$n^{(2)}$	$n^{(2)}$	$n^{(3)}$	$n^{(4)}$
		μ_4	$\mu_{3,1}$	$\mu_{2,2}$	$\mu_{2,1^2}$	μ_{1^4}
$\mu_{1:m_4}$	n^{-1}	1				
$\mu_{11:m_1m_3}$	n^{-2}	1	1			
$\mu_{2:m_2}$	n^{-2}	1		1		
$\mu_{21:m_1m_2}$	n^{-3}	1	2	1	1	
$\mu_{4:m_1}$	n^{-4}	1	4	3	6	1

(6)												
	Coef.	n	$n^{(2)}$	$n^{(2)}$	$n^{(2)}$	$n^{(3)}$	$n^{(3)}$	$n^{(3)}$	$n^{(4)}$	$n^{(4)}$	$n^{(5)}$	$n^{(6)}$
		μ_6	$\mu_{5,1}$	$\mu_{4,2}$	$\mu_{3,3}$	$\mu_{4,1^2}$	$\mu_{3,2,1}$	μ_{2^3}	$\mu_{3,1^3}$	$\mu_{2^2,1^2}$	$\mu_{2,1^4}$	μ_{1^6}
$\mu_{1:m_6}$	n^{-1}	1										
$\mu_{11:m_1m_5}$	n^{-2}	1	1									
$\mu_{11:m_2m_4}$	n^{-2}	1		1								
$\mu_{2:m_3}$	n^{-3}	1			1							
$\mu_{21:m_1m_4}$	n^{-3}	1	2	1		1						
$\mu_{111:m_1m_2m_3}$	n^{-3}	1	1	1	1		1					
$\mu_{3:m_2}$	n^{-3}	1		3				1				
$\mu_{31:m_1m_3}$	n^{-4}	1	3	3	1	3	3		1			
$\mu_{22:m_1m_2}$	n^{-4}	1	2	3	2	1	4	1		1		
$\mu_{41:m_1m_3}$	n^{-5}	1	4	7	4	6	16	3	4	6	1	
$\mu_{6:m_1}$	n^{-6}	1	6	15	10	15	60	15	20	45	15	1

After writing $(x_i - m_i)^t$ as the sum of the general term of a binomial series and then expanding the resulting right member of (1.21) as a product of power sums [20; p. 19], we get

$$(1.22) \quad \mu_{s;\bar{m}_t} = \sum \frac{s!}{r_1! r_2! \dots r_v! \pi_1! \pi_2! \dots} \sum_{\substack{i_1, i_2, \dots, i_v=0 \\ i_1 \neq i_2 \neq \dots \neq i_v}}^t (-1)^\rho \binom{t}{i_1}^{r_1} \binom{t}{i_2}^{r_2} \dots \binom{t}{i_v}^{r_v} \mu_{r_1 r_2 \dots r_v; m_t - i_1 m_t - i_2 \dots m_t - i_v m_1}$$

where $\sum_{j=1}^v r_j = s$, $\sum_{j=1}^v i_j r_j = \rho$ and π_1, π_2, \dots are the numbers of the repeated parts of s .

The mean of the t th central moment takes the following simple form,

$$(1.23) \quad \mu_{1;\bar{m}_t} = \sum_{i=0}^t (-1)^i \binom{t}{i} \mu_{1i; m_t - i m_1}$$

where the moments in the right member of (1.23) through weight six are given in the tables of section four. Also,

$$(1.24) \quad \mu_{2;\bar{m}_2} = \mu_{2;m_2} - 2' \mu_{21;m_1 m_2} + \mu_{4;m_1}.$$

$$(1.25) \quad \mu_{3;\bar{m}_2} = \mu_{3;m_2} - 3' \mu_{22;m_1 m_2} + 3' \mu_{41;m_1 m_2} - \mu_{6;m_1}.$$

$$(1.26) \quad \begin{aligned} \mu_{2;\bar{m}_3} = & \mu_{2;m_3} + 9' \mu_{22;m_1 m_2} + 4' \mu_{6;m_1} - 6' \mu_{111;m_1 m_2 m_3} \\ & + 4' \mu_{31;m_1 m_3} - 12' \mu_{41;m_1 m_2}. \end{aligned}$$

After substituting from the tables of section four, (1.23) through (1.26) become

$$(1.27) \quad \mu_{1;\bar{m}_2} = \frac{n^{(2)}}{n^2} [\mu_2 - \mu_{1,1}].$$

$$(1.28) \quad \mu_{1;\bar{m}_3} = \frac{n^{(3)}}{n^3} [\mu_3 - 3\mu_{2,1} + 2\mu_{1,1}].$$

$$(1.29) \quad \begin{aligned} \mu_{1;\bar{m}_4} = & \frac{1}{n^4} [n^{(2)}(n^2 - 3n + 3)(\mu_4 - 4\mu_{3,1}) + 3n^{(2)}(2n - 3)\mu_{2,2} \\ & + 3n^{(4)}(2\mu_{2,1^2} - \mu_{1,1^4})]. \end{aligned}$$

$$(1.30) \quad \begin{aligned} \mu_{1;\bar{m}_5} = & \frac{1}{n^5} [n^{(3)}(n^2 - 2n + 2)(\mu_5 - 5\mu_{4,1}) + 10n^{(3)}(n - 2)\mu_{3,2} \\ & + 10n^{(3)}(n + 1)(n - 4)\mu_{3,1^2} - 30n^{(3)}(n - 2)\mu_{2^2,1} \\ & - 10n^{(4)}(3n - 4)\mu_{2,1^3} + 4n^{(5)}\mu_{1,1^5}]. \end{aligned}$$

$$\begin{aligned}
 {}'\mu_{1:\bar{m}_6} &= \frac{1}{n^6} [n^2(n^4 - 5n^3 + 10n^2 - 10n + 5)(\mu_6 - \mu_{6,1}) \\
 &\quad + 15n^{(2)}(n^3 - 4n^2 + 7n - 5)\mu_{4,2} - 10n^{(3)}(2n^2 - 6n + 5)\mu_{3,1} \\
 &\quad + 15n^{(3)}(n^3 - 4n^2 + 6n - 5)\mu_{4,1,1} - 60n^{(3)}(n^2 - 4n + 5)\mu_{3,2,1} \\
 &\quad + 15n^{(3)}(3n - 5)\mu_{2,3} - 20n^{(4)}(n^2 - 3n + 5)\mu_{3,1,1} \\
 &\quad + 45n^{(4)}(2n - 5)\mu_{2,2,1,1} + 15n^{(5)}(n - 5)\mu_{2,1,1} - 5n^{(6)}\mu_{1,6}].
 \end{aligned}
 \tag{1.31}$$

$${}'\mu_{2:\bar{m}_2} = \frac{1}{n^4} [n^{(2)}(n - 1)(\mu_4 - 4\mu_{3,1}) + n^{(3)}(n + 1)\mu_{2,2} - n^{(4)}(2\mu_{2,1,1} - \mu_{1,4})].
 \tag{1.32}$$

$$\begin{aligned}
 {}'\mu_{3:\bar{m}_2} &= \frac{1}{n^6} [n^{(2)}(n - 1)^2(\mu_6 - 6\mu_{5,1}) + 3n^{(2)}(n - 1)(n^2 - 2n + 5)\mu_{4,2} \\
 &\quad - 2n^{(2)}(3n^2 - 6n + 5)\mu_{3,3} + n^{(3)}(n^3 - 3n^2 + 9n - 15)\mu_{2,1} \\
 &\quad - 3n^{(3)}(n - 1)(n - 5)\mu_{4,1,1} - 12n^{(3)}(n^2 - 4n + 5)\mu_{3,2,1} \\
 &\quad + 4n^{(4)}(3n - 5)\mu_{3,1,1} - 3n^{(4)}(n^2 - 6n + 15)\mu_{2,2,1,1} \\
 &\quad + n^{(6)}(3\mu_{2,1,1} - \mu_{1,6})].
 \end{aligned}
 \tag{1.33}$$

$$\begin{aligned}
 {}'\mu_{2:\bar{m}_3} &= \frac{1}{n^6} [n^{(2)}(n - 1)^2(n - 2)(\mu_6 - 6\mu_{5,1}) - 3n^{(2)}(n - 2)^2(2n - 5)\mu_{4,2} \\
 &\quad + n^{(2)}(n - 2)^2(n^2 - 2n + 10)\mu_{3,3} \\
 &\quad - 6n^{(3)}(n - 2)(n^2 - 6n + 20)\mu_{3,2,1} + 3n^{(3)}(n - 2)(7n - 10)\mu_{4,1,1} \\
 &\quad + 3n^{(3)}(3n^2 - 12n + 20)\mu_{2,3} + 4n^{(4)}(n - 2)(n - 10)\mu_{3,1,1} \\
 &\quad + 9n^{(4)}(n^2 - 8n + 20)\mu_{2,2,1,1} - 4n^{(6)}(3\mu_{2,1,1} - \mu_{1,6})].
 \end{aligned}
 \tag{1.34}$$

6. The variance of the variance of samples of n . The variance of the variance of samples of n , when the moments of the universe are measured about a fixed point, is defined as

$${}'\bar{\mu}_{2:\bar{m}_2} = {}'\mu_{2:\bar{m}_2} - [{}'\mu_{1:\bar{m}_2}]^2.
 \tag{1.35}$$

Therefore, from (1.27) and (1.32),

$$\begin{aligned}
 {}'\mu_{2:\bar{m}_2} &= \frac{1}{n^4} [n^{(2)}(n - 1)(\mu_4 - 4\mu_{3,1}) + n^{(3)}(n - 1)\mu_{2,2} - n^{(4)}(2\mu_{2,1,1} - \mu_{1,4})] \\
 &\quad - \left(\frac{n - 1}{n} \right)^2 (\mu_2 - \mu_{1,1})^2.
 \end{aligned}
 \tag{1.36}$$

Tchouproff [4; p. 492] gave a formula (8) for the variance of the sample variance but his result is unwieldy due to the fact that moments of the universe are measured about the mean.

7. Conventional infinite sampling formulas derived from generalized sampling formulas. The term "infinite sampling" is to be interpreted as meaning: *sampling from an unlimited supply or sampling from a limited supply with repetitions permitted*. In each of these situations the variates are independent [5; p. 79].

First, it is assured that the n populations are identical, that is, ${}_1X = {}_2X = \dots = {}_nX$. This assumption results in the fact that, for a fixed t , ${}_1\mu_t = {}_2\mu_t = \dots = {}_n\mu_t$ and ${}_1\bar{\mu}_t = {}_2\bar{\mu}_t = \dots = {}_n\bar{\mu}_t$. Therefore, under the assumption of identical populations, every moment may be interpreted as either the moment of n identical populations or as the moment of a single population. The only other assumption is that the sampling is "infinite".

From the condition of independence [3; p. 141], we have

$$E(r_1 x_{i_{r_1}}^{t_1} r_2 x_{i_{r_2}}^{t_2} \dots r_v x_{i_{r_v}}^{t_v}) = (E_{r_1} x_{i_{r_1}}^{t_1})(E_{r_2} x_{i_{r_2}}^{t_2}) \dots (E_{r_v} x_{i_{r_v}}^{t_v}).$$

Therefore,

$$r_1 r_2 \dots r_v \mu_{t_1 t_2 \dots t_v} = r_1 \mu_{t_1} r_2 \mu_{t_2} \dots r_v \mu_{t_v}.$$

Combining the condition of independence with that of identical populations, we have

$$(1.37) \quad \frac{1}{n^{(v)}} S_v r_1 r_2 \dots r_v \mu_{t_1 t_2 \dots t_v} = \frac{1}{n^{(v)}} S_v r_1 \mu_{t_1} r_2 \mu_{t_2} \dots r_v \mu_{t_v} = \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}.$$

By (1.16) and (1.37), we may write

$$(1.38) \quad \mu_{t_1 t_2 \dots t_v} = \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}.$$

Since the only terms of the generalized sampling formulas are affected by the assumption of "infinite sampling" are those of the form $\mu_{t_1 t_2 \dots t_v}$, the problem of obtaining conventional infinite sampling formulas from generalized sampling formulas is, in practice, a mechanical one. Simply write terms of the form $\mu_{t_1 t_2 \dots t_v}$ which appear in a generalized sampling formula, as $\mu_{t_1} \mu_{t_2} \dots \mu_{t_v}$ and one automatically obtains the corresponding infinite sampling formula.

As an illustration of the method, consider the generalized sampling formula (1.36) for the variance of the sample variance. When (1.38) is utilized to change it into the corresponding infinite sampling formula, (1.36) becomes

$$(1.39) \quad \bar{\mu}_{2;\bar{m}_2} = \frac{n^{(2)}}{n^4} [(n-1)(\mu_4 - 4\mu_3\mu_1) - (n-3)\mu_2^2 + 2(2n-3)(2\mu_2\mu_1^2 - \mu_1^4)],$$

which is the usual formula [20; p. 75] for the variance of the sample variance when the moments of the universe are measured about a fixed point. If it is assumed that the moments of ${}_nU_N$ are measured about the mean, formula (1.39) becomes

$$(1.40) \quad \bar{\mu}_{2;\bar{m}_2} = \frac{n^{(2)}}{n^4} [(n-1)\bar{\mu}_4 - (n-3)\bar{\mu}_2^2],$$

which was published by "Student" [1; p. 3] in 1908.

8. Conventional finite sampling formulas derived from generalized sampling formulas. The term "finite sampling" is to be interpreted as meaning: *sampling from a limited supply when repetitions are not permitted.*

In order to reduce generalized sampling formulas to the corresponding formulas for finite sampling, the assumptions are made that the n populations are identical and that N and n are finite, $N > n$. The selection of variates which enter each sample is restricted in the following manner. If a variate having a given post-subscript is chosen, then no other variate having the same post-subscript may be chosen for the same sample.

Now it is evident that terms of the form $\mu_{t_1 t_2 \dots t_v}$ must be redefined on the basis of the preceding assumptions. From the expansions [20; p. 32] of power product sums in terms of products of power sums, we get the formulas for $\mu_{t_1 t_2 \dots t_v}$ which are given in the following tables.

The formulas in the tables of this section are called *transformation formulas for finite sampling* or more briefly *transformation formulas*.

The transformation of generalized sampling formulas into corresponding finite sampling formulas is illustrated by the substitution of $\frac{N^2 \mu_1^2 - N \mu_2}{N^{(2)}}$ for $\mu_{1,1}$ in (1.27). We get

$$(1.41) \quad \mu_{1,1} = \frac{N(n-1)}{n(N-1)} [\mu_2 - \mu_1^2],$$

which is the well-known finite sampling formula for the mean of the variance of samples of n .

From this and the preceding section it is evident that the generalized sampling formulas may be considered as formulas for either infinite or finite sampling depending upon the interpretation given to terms of the form $\mu_{t_1 t_2 \dots t_v}$.

9. Transformation of infinite sampling formulas into corresponding finite sampling formulas. It is a well-known fact that infinite sampling formulas may be obtained from those for finite sampling by letting the size of the parent population become infinite. But, prior to this paper, apparently no one has presented a method of obtaining finite sampling formulas from infinite sampling formulas. However, by making use of the relations between finite, infinite, and generalized sampling, we shall demonstrate that it is possible to transform any infinite sampling formula into the corresponding finite sampling formula.

Since the infinite sampling formulas are obtained from the generalized sampling formulas by replacing

$$\mu_{t_1 t_2 \dots t_v} \text{ by } \mu_{t_1} \mu_{t_2} \dots \mu_{t_v}$$

it follows that generalized sampling formulas may be obtained from the infinite

TABLE II

(1)			(2)				(3)				
	Coef.	N		Coef.	N	N^2		Coef.	N	N^2	N^3
		μ_1			μ_2	μ_1^2			μ_3	$\mu_2\mu_1$	μ_1^3
μ_1	$N^{(-1)}$	1	μ_2	$N^{(-1)}$	1		μ_3	$N^{(-1)}$	1		
			$\mu_{1,1}$	$N^{(-2)}$	-1	1	$\mu_{2,1}$	$N^{(-2)}$	-1	1	
							μ_{1^2}	$N^{(-3)}$	2	-3	1

(5)									
	Coef.	N	N^2	N^2	N^3	N^3	N^4	N^5	
		μ_5	$\mu_4\mu_1$	$\mu_3\mu_2$	$\mu_2\mu_1^2$	$\mu_2^2\mu_1$	$\mu_2\mu_1^3$	μ_1^5	
μ_5	$N^{(-1)}$	1							
$\mu_{4,1}$	$N^{(-2)}$	-1	1						
$\mu_{3,2}$	$N^{(-2)}$	-1		1					
$\mu_{3,1^2}$	$N^{(-3)}$	2	-2	-1	1				
$\mu_{2^2,1}$	$N^{(-3)}$	2	-1	-2		1			
$\mu_{2,1^3}$	$N^{(-4)}$	-6	6	5	-3	-3	1		
μ_{1^5}	$N^{(-5)}$	24	-30	-20	20	15	-10	1	

(4)						
	Coef.	N	N^2	N^2	N^3	N^4
		μ_4	$\mu_{3,1}$	μ_2^2	$\mu_2\mu_1^2$	μ_1^4
μ_4	$N^{(-1)}$	1				
$\mu_{3,1}$	$N^{(-2)}$	-1	1			
$\mu_{2,2}$	$N^{(-2)}$	-1		1		
$\mu_{2,1^2}$	$N^{(-3)}$	2	-2	-1	1	
μ_{1^4}	$N^{(-4)}$	-6	8	3	-6	1

(6)												
	Coef.	N	N^2	N^2	N^2	N^3	N^3	N^3	N^4	N^4	N^5	N^6
		μ_6	$\mu_5\mu_1$	$\mu_4\mu_2$	μ_3^2	$\mu_4\mu_1^2$	$\mu_3\mu_2\mu_1$	μ_2^3	$\mu_3\mu_1^3$	$\mu_2^2\mu_1^2$	$\mu_2\mu_1^4$	μ_1^6
μ_6	$N^{(-1)}$	1										
$\mu_{5,1}$	$N^{(-2)}$	-1	1									
$\mu_{4,2}$	$N^{(-2)}$	-1		1								
$\mu_{3,3}$	$N^{(-2)}$	-1			1							
$\mu_{4,1^2}$	$N^{(-3)}$	2	-2	-1		1						
$\mu_{3,2,1}$	$N^{(-3)}$	2	-1	-1	-1		1					
μ_{2^3}	$N^{(-3)}$	2		-3				1				
$\mu_{3,1^3}$	$N^{(-4)}$	-6	6	3	2	-3	-3		1			
$\mu_{2^2,1^2}$	$N^{(-4)}$	-6	4	5	2	-1	-4	-1		1		
$\mu_{2,1^4}$	$N^{(-5)}$	24	-24	-18	-8	12	20	3	-4	-6	1	
μ_{1^6}	$N^{(-6)}$	-120	144	90	40	-90	-120	-15	40	45	-15	1

formulas by replacing

$$(1.42) \quad \mu_{t_1} \mu_{t_2} \cdots \mu_{t_v} \text{ by } \mu_{t_1 t_2 \cdots t_v}.$$

However, it must be emphasized that the application of (1.42) demands formulae which are expressed in terms of moments of sample moments rather than central moments of sample moments (although the sample moments may be measured about a fixed point or about the mean) and the moments of the universe must be measured about a fixed point. The reason for these restrictions is to insure that each term is accounted for individually.

After replacements (1.42) are made in the formula for sampling from an infinite population, the resulting formula is the corresponding generalized one. The step to the corresponding finite sampling formulas is simply the one outlined in section eight, namely, the use of the transformation formulas.

We shall consider, as the first illustration, the infinite sampling formula for the mean of the sample variance when the moments of the parent population are measured about the mean. The formula is

$$(1.43) \quad \mu_{1:\bar{m}_2} = \frac{n-1}{n} \bar{\mu}_2.$$

When (1.43) is expressed in terms of moments of the parent population about a fixed point, we have

$$(1.44) \quad \mu'_{1:\bar{m}_2} = \frac{n-1}{n} [\mu_2 - \mu_1^2].$$

Following (1.42), μ_1^2 is replaced by $\mu_{1,1}$ and (1.44) becomes (1.27). The use of the transformation formula for $\mu_{1,1}$ gives (1.41) which, when the moments of the parent population are measured about the mean, becomes

$$(1.45) \quad \mu_{1:\bar{m}_2} = \frac{N(n-1)}{n(N-1)} \bar{\mu}_2.$$

Infinite sampling formulas expressed in terms of moment-function, may be similarly transformed into the corresponding finite sampling formulas. For example, Craig [9; p. 57] gives the second Thiele seminvariant of the variance of samples as

$$(1.46) \quad \lambda_{2:\bar{m}_2} = \frac{(n-1)^2}{n^3} \lambda_4 + 2 \frac{(n-1)}{n^2} \lambda_2^2.$$

First, we express (1.46) in terms of moments about a fixed point by use of the formulas relating Thiele seminvariants and moments [9; p. 12]. We also recall that the resulting formula should be expressed in terms of moments of sample moments rather than in terms of central moments of sample moments. We obtain

$$(1.47) \quad \mu'_{2:\bar{m}_2} = \frac{(n-1)}{n^3} [(n-1)\mu_4 - 4(n-1)\mu_3\mu_1 + (n^2 - 2n + 3)\mu_2^2 \\ - 2(n-2)(n-3)\mu_2\mu_1^2 + (n-2)(n-3)\mu_1^4].$$

The next step is to transform (1.47) into the corresponding generalized sampling formula by use of (1.42). We obtain (1.32). Since we desire to obtain the finite sampling formula which exactly corresponds to (1.46), it is necessary to transform (1.32) from the second moment of \bar{m}_2 to the variance of \bar{m}_2 and we get (1.36). Next the transformation formulas are applied to (1.36). When the moments of the parent population are measured about the mean and are replaced by Thiele seminvariants, (1.36) becomes

$$(1.48) \quad \lambda_{2:\bar{m}_2} = \frac{N(N-n)(n-1)}{n^3(N-1)^2(N-2)(N-3)} [(N-1)(Nn - N - n - 1)\lambda_4 \\ + 2(N^2n - 3Nn - 3N + 3n + 3)\lambda_2^2].$$

Formula (1.48) gives the second Thiele seminvariant of the variance of samples of n drawn from a finite parent population of N . When $N \rightarrow \infty$, in (1.48), we obtain immediately (1.46).

It is generally true that infinite sampling formulas are more easily derived than are the corresponding finite sampling formulas. The methods of this section make it possible to derive the desired sampling formulas for the infinite parent population and then transform these infinite sampling formulas into the corresponding finite sampling formulas.

II. MOMENT FUNCTION ADJUSTMENTS FOR GROUPED DATA

A given distribution of discrete variates may be grouped in " k groupings of k ". We desire to find the correction which eliminates the error made in replacing a given moment of the original distribution by the average of the corresponding moments of the k grouped-distributions.

Formulas for the adjustments for moments of a grouped-distribution of discrete variates were first given (without proof) in the Editorial of Vol. I, No. 1 of the *Annals of Mathematical Statistics*. Later, more satisfactory derivations of adjustment formulas were given by Abernethy [24] Craig [25] and Carver [26]. However, it was observed by Carver [26; p. 162] that the developments of Abernethy and Craig are adjustments about a fixed point and that they fail to hold for the case of expectations of central moments if we accept the definition

$$\mu_{1:\bar{\mu}_t} = \frac{1}{k} \sum_{r=1}^k r\bar{\mu}_t, \quad (t = 2, 3, \dots).$$

Here $r\bar{\mu}_t$ represents the t th central moment of the r th grouped-distribution. The formula for the true value of $\mu_{1:\bar{\mu}_2}$ was supplied by Carver [26; p. 162] but he did not indicate a general method which might be used for the derivation of $\mu_{1:\bar{\mu}_t}$, ($t > 2$).

A distribution of discrete variates grouped in " k groupings of k " is a special case of a universe of n finite populations and hence the methods and formulas for the expectations of population moments are applicable to our present problem.

It is found that the adjustment formulas for moment-functions of grouped data involve central moments of a rectangular distribution. It will be convenient for our present purposes to give a brief treatment of the moment-functions of a rectangular distribution.

1. Moment-functions of a rectangular distribution. Consider the rectangular distribution of discrete variates,

$$(2.1) \quad h, 2h, 3h, \dots, kh.$$

It is readily shown that the moment generating function of (2.1),

$$(2.2) \quad G_x(\theta) = \mu_0 + \mu_1\theta + \mu_2\frac{\theta^2}{2!} + \dots + \mu_n\frac{\theta^n}{n!} + \dots$$

may be written

$$(2.3) \quad G_x(\theta) = \frac{e^{\frac{1}{2}(k+1)h\theta} \sinh \frac{1}{2}kh\theta}{k \sinh \frac{1}{2}h\theta}.$$

Setting the expansion of the right member of (2.3) equal to the right member of (2.2) and equating coefficients of like powers of θ , we obtain the following recursion formula for the moments of (1.1)

$$(2.4) \quad \frac{(n+1)^{(1)}}{1!} \mu_{n:R} - \frac{(n+1)^{(2)}}{2!} h \mu_{n-1:R} + \dots \\ + (-1)^{r-1} \frac{(n+1)^{(r)}}{r!} h^{r-1} \mu_{n-r+1:R} + \dots = k^n h^n,$$

where $\mu_{n:R}$ represents the n th moment of a rectangular distribution. Formulas for $\mu_{n:R}$, ($n = 0, 1, \dots, 10$) are given below. See Sasuly [27; p. 27].

$$\begin{aligned} \mu_{0:R} &= 1. \\ \mu_{1:R} &= \frac{1}{2}(k+1)h. \\ \mu_{2:R} &= \frac{1}{6}(k+1)(2k+1)h^2 = \frac{1}{3}(2k+1)h \mu_{1:R}. \\ \mu_{3:R} &= \frac{1}{4}(k+1)^2 kh^3 = kh \mu_{1:R}^2. \\ \mu_{4:R} &= \frac{1}{5}(3k^2 + 3k - 1)h^2 \mu_{2:R}. \\ (2.5) \quad \mu_{5:R} &= \frac{1}{3}(2k^2 + 2k + 1)h^2 \mu_{3:R}. \\ \mu_{6:R} &= \frac{1}{7}(3k^4 + 6k^3 - 3k + 1)h^4 \mu_{2:R}. \\ \mu_{7:R} &= \frac{1}{6}(3k^4 + 6k^3 - k^2 - 4k + 2)h^4 \mu_{3:R}. \\ \mu_{8:R} &= \frac{1}{18}(5k^6 + 15k^5 + 5k^4 - 15k^3 - k^2 + 9k - 3)h^6 \mu_{2:R}. \\ \mu_{9:R} &= \frac{1}{5}(2k^6 + 6k^5 + k^4 - 8k^3 + k^2 + 6k - 3)h^6 \mu_{3:R}. \\ \mu_{10:R} &= \frac{1}{11}(3k^8 + 12k^7 + 8k^6 - 18k^5 - 10k^4 + 24k^3 + 2k^2 - 15k + 5)h^8 \mu_{2:R}. \end{aligned}$$

The deviations about the mean of (2.1) are

$$(2.6) \quad -\frac{1}{2}(k-1)h, \quad -\frac{1}{2}(k-3)h, \dots, \frac{1}{2}(k-3)h, \quad \frac{1}{2}(k-1)h.$$

Therefore,

$$(2.7) \quad \bar{\mu}_{2n+1;R} = 0.$$

If we denote (2.6) by \bar{x} , we have

$$(2.8) \quad G_{\bar{x}}(\theta) = \frac{\sinh \frac{1}{2}(\bar{x}h\theta)}{\bar{x} \sinh \frac{1}{2}(\bar{x}h\theta)}.$$

The recursion formula for central moments of (2.1) is

$$(2.9) \quad \frac{(2n+1)^{(1)}}{1!} \bar{\mu}_{2n;R} + \frac{h^2 (2n+1)^{(3)}}{2^2 3!} \bar{\mu}_{2n-2;R} + \dots + \frac{h^r (2n+1)^{(r+1)}}{2^r (r+1)!} \bar{\mu}_{2n-r;R} + \dots = \frac{k^{2n} h^{2n}}{2^{2n}}.$$

Formulas for $\bar{\mu}_{2n;R}$, ($n = 0, 1, \dots, 5$) are given below. See [27; p. 27].

$$(2.10) \quad \begin{aligned} \bar{\mu}_{0;R} &= 1, \\ \bar{\mu}_{2;R} &= \frac{1}{12}(k^2 - 1)h^2, \\ \bar{\mu}_{4;R} &= \frac{1}{24}(3k^2 - 7)h^2 \bar{\mu}_{2;R}, \\ \bar{\mu}_{6;R} &= \frac{1}{112}(3k^4 - 18k^2 + 31)h^4 \bar{\mu}_{2;R}, \\ \bar{\mu}_{8;R} &= \frac{1}{640}(5k^6 - 55k^4 + 239k^2 - 381)h^6 \bar{\mu}_{2;R}, \\ \bar{\mu}_{10;R} &= \frac{1}{2816}(3k^8 - 52k^6 + 410k^4 - 1636k^2 + 2555)h^8 \bar{\mu}_{2;R}. \end{aligned}$$

From the relation which connects Thiele seminvariants and the moment generating function, we get, see [25; p. 57],

$$(2.11) \quad \begin{aligned} \lambda_{0;R} &= 0, \quad \lambda_{1;R} = \frac{(k+1)h}{2}, \quad \lambda_{2n+1;R} = 0, \\ \lambda_{2n;R} &= (-1)^{n+1} \frac{B_n h^{2n} (k^{2n} - 1)}{2n}, \quad n = 1, 2, 3, \dots \end{aligned}$$

where $\lambda_{n;R}$ represents the n th Thiele seminvariant of a rectangular distribution of discrete variates and B_n , ($n = 1, 2, \dots$), the Bernoulli numbers: $\frac{1}{6}, \frac{1}{30}, \dots$.

In each of the cases considered in this section, corresponding formulas may be found for a rectangular distribution of continuous variates by setting $h = m/k$ (which makes the range m with k subdivisions) and then letting $k \rightarrow \infty$.

2. Adjustments for moments. As our basic distribution we consider the set of discrete variates, x_i , ($i = 1, 2, \dots, N$), where some of the x_i 's may not be distinct. We assume that the given distribution is grouped in " k groupings of k ".

When x_i is placed in the r th position of a class, the limits of the class are $x_i - (r - 1)h$ and $x_i + (k - r)h$ and the class mark is $x_i + \left[\frac{k - (2r - 1)}{2} \right] h$. Thus, when the class mark is used as the value of x_i , the quantity $\left[\frac{k - (2r - 1)}{2} \right] h$ is added to the true value of x_i . Therefore, when the expected value of a particular moment for " k groupings of k " is found, each variate has made a definite contribution as it was placed in each of the k positions of a class.

For convenience, we define

$$(2.12) \quad e_r = \left[\frac{k - (2r - 1)}{2} \right] h, \quad (r = 1, 2, \dots, k).$$

As was previously indicated, the expected value of a given moment involves the contribution of each variate as it occupies the k class positions. A convenient method of finding these contributions is by means of a universe ${}_kU_N$ which is composed of the populations ${}_rX$, ($r = 1, 2, \dots, k$). The r th population consists of the values of the variates when they occupy the r th position of the class. Hence ${}_rX$ consists of ${}_rx_i = x_i + e_r$, ($i = 1, 2, \dots, N$).

The notation for moments is the same as that of Part I. Since ${}_kU_N$ is of the same form as the universe studied in Part I, we use the definitions (1.1) of that part.

The expected value of the t th moment is

$$\begin{aligned} \mu_{1:\mu_t} &= \frac{1}{k} \sum_{r=1}^k E(x_i + e_r)^t \\ &= \sum_{s=0}^t \binom{t}{s} \left[\frac{1}{k} \sum_{r=1}^k e_r^s \right] \mu_{t-s}. \end{aligned}$$

Many devices have been used by previous writers [24; p. 269], [25; p. 57], [26; p. 157], to evaluate terms of the form $\frac{1}{k} \sum_{r=1}^k e_r^s$. However, it should be noticed that the quantities e_r , ($i = 1, 2, \dots, k$), are respectively identical with the deviations (2.6) about the mean of a rectangular distribution of discrete variates. It follows that

$$\bar{\mu}_{s;R} = \frac{1}{k} \sum_{r=1}^k e_r^s.$$

And since $\bar{\mu}_{2s+1;R} = 0$, we have

$$(2.13) \quad \mu_{1:\mu_t} = \sum_{s=0}^{[t/2]} \binom{t}{2s} \mu_{t-2s} \bar{\mu}_{2s;R}.$$

Formulas for $\bar{\mu}_{2s;R}$, ($s = 0, 1, \dots, 5$) are given by (2.10).

If the class marks are selected as the unit of x , we set $h = 1$ in (2.10). If the

class interval is chosen as the unit of x , we set $h = 1/k$ in (2.10). If k consecutive values of the discrete variable are grouped in a frequency class of width m , we put $h = m/k$ in (2.10).

Usually we desire to estimate the value of the moments that would have been obtained if we had not grouped the data. Therefore (2.13) is solved for the moments of the ungrouped data. We have

$$(2.14) \quad \mu_t = \sum_{s=0}^{\lfloor t/2 \rfloor} \binom{t}{2s} P_{2s} \mu_{1:\mu_t-2s}$$

wherein

$$P_{2s} = \sum \frac{(-1)^s (2s)! \rho! \bar{\mu}_{2p_1:R}^{\pi_1} \bar{\mu}_{2p_2:R}^{\pi_2} \cdots \bar{\mu}_{2p_v:R}^{\pi_v}}{[(2p_1)!]^{\pi_1} [(2p_2)!]^{\pi_2} \cdots [(2p_v)!]^{\pi_v} \pi_1! \pi_2! \cdots \pi_v!},$$

the summation being taken for every possible product of moments for which

$$\sum_{i=1}^v p_i = s, \quad \sum_{i=1}^v \pi_i = \rho.$$

Formulas, corresponding to (2.13) and (2.14), for a distribution of continuous variates are written by replacing the moment symbols for discrete variates by those for continuous variates.

3. Adjustments for central moments. Consider the universe U which consists of the population ${}_rX$, ($r = 1, 2, \dots, k$), where ${}_rX$ is the r th grouped-distribution.

The expected value of the t th central moment of the k grouped-distribution is given by (1.3), (1.4) and (1.5) of Part I, where now $\mu_{1:\mu_t-i}$ is given by (2.13) of the preceding section. Thus, the development of this section is identical with that of section one of Part I with the single exception that $\mu_{1:\mu_t} = \mu_t$ no longer holds but is replaced by $\mu_{1:\mu_t} = \mu_t + \text{a correction}$. Therefore, the formulas for the adjustments for central moments may be obtained immediately from the formulas derived in section one, Part I, if the corrections of the preceding section are inserted. We have

$$(2.15) \quad \mu_{1:\bar{\mu}_2} = \bar{\mu}_2 + \bar{\mu}_{2:R} - \bar{\mu}_{2:\mu_1}$$

$$(2.16) \quad \mu_{1:\bar{\mu}_3} = \bar{\mu}_3 + 6\bar{\mu}_1\bar{\mu}_{2:\mu_1} - 3\bar{\mu}_{11:\mu_1\mu_2} + 2\bar{\mu}_{3:\mu_1}$$

$$(2.17) \quad \begin{aligned} \mu_{1:\bar{\mu}_4} = & \bar{\mu}_4 + 6\bar{\mu}_2\bar{\mu}_{2:R} + \bar{\mu}_{4:R} + 6(\bar{\mu}_2 - 2\bar{\mu}_1^2 + \bar{\mu}_{2:R})\bar{\mu}_{2:\mu_1} \\ & + 12\bar{\mu}_1\bar{\mu}_{11:\mu_1\mu_2} - 12\bar{\mu}_1\bar{\mu}_{3:\mu_1} - 4\bar{\mu}_{11:\mu_1\mu_3} \\ & + 6\bar{\mu}_{21:\mu_1\mu_2} - 3\bar{\mu}_{4:\mu_1} \end{aligned}$$

The moments of the ungrouped data can be obtained readily from formulas (2.15) through (2.17).

Adjustment formulas for central moments of a distribution of continuous variates may be obtained from (2.13) by replacing the moment symbols for

discrete variates by those for continuous variates and taking the moments about the mean. Also, it may be observed that adjustment formulas for central moments of a distribution of continuous variates may be obtained from formulas (1.3), (1.4) and (1.5) of Part I, provided the moment symbols are exchanged as indicated above and terms of the form $\bar{\nu}_{s_1 s_2 \dots s_r; \nu_{t_1} \nu_{t_2} \dots \nu_{t_r}}$ are set equal to zero.

4. Usual adjustments for Thiele seminvariants. The usual adjustments for Thiele seminvariants, for the univariate discrete population, may be developed directly by use of one of the fundamental properties of Thiele seminvariants.

It is assumed (see [25; p. 55]) that k consecutive values of the discrete variable are grouped in a frequency class of width m . The k smaller intervals of width $m/k = h$ go to make up the class width m , the actual points representing the k values of the variable being plotted at the centers of the sub-intervals. Now, let us suppose that each of the k consecutive boundary points of the subintervals is as likely to be chosen as a boundary point of the larger intervals as any other. Then, if x_i is the class mark of the i th frequency class, for any true value, x , of the discrete variable included in this frequency class, we have

$$x_i = x + e_r$$

in which x and e_r are independent variables and e_r takes on the k values (2.12) with equal relative frequencies $1/k$.

Since we have noted that the equally likely values which e_r may take on are deviations about the mean of a rectangular distribution of discrete variates, we employ the cumulative property of Thiele seminvariants [9; p. 4] and obtain directly

$$(2.18) \quad \lambda'_{t;x} = \lambda_{t;x} + \lambda_{t;R}, \quad (t = 1, 2, \dots),$$

where $\lambda'_{t;x}$ is the t th seminvariant computed from the grouped data, $\lambda_{t;x}$ is the t th seminvariant computed from the ungrouped data and $\lambda_{t;R}$ is defined by (2.11).

Formulas corresponding to (2.18), for special values of t , are given by Craig [25; p. 57]. However, the present development indicates the dependence of adjustment formulas on central moments of a rectangular distribution and provides a general formula for these adjustments which is expressed completely in terms of Thiele seminvariants.

5. New adjustments for Thiele seminvariants. If we accept the definition

$$\mu_{1;\bar{\mu}_t} = \frac{1}{k} \sum_{r=1}^k r \bar{\mu}_t, \quad (t = 2, 3, \dots),$$

then (2.18) is at best only an approximation formula. We now desire exact formulas for $\mu_{1;\lambda_t}$ for the case of a grouped-distribution of discrete variates.

First (1.9) is used and terms of the form $\mu_{s_1 s_2 \dots s_r; \mu_1 \mu_2 \dots \mu_r}$ are evaluated in terms of central moments by (1.3). Then terms of the form $\mu_{1; \mu_t}$ are evaluated by (2.13) and finally the relations between moments and Thiele seminvariants are employed. Exact formulas for the expected values of the second, third, and fourth Thiele seminvariants for grouped-distributions of discrete variables are given below.

$$(2.19) \quad \mu_{1; \lambda_2} = \lambda_2 + \lambda_{2; R} - \bar{\mu}_{2; \mu_1}.$$

$$(2.20) \quad \mu_{1; \lambda_3} = \lambda_3 + 6\lambda_1 \bar{\mu}_{2; \mu_1} - 3\bar{\mu}_{11; \mu_1 \mu_2} + 2\bar{\mu}_{3; \mu_1}.$$

$$(2.21) \quad \begin{aligned} \mu_{1; \lambda_4} = & \lambda_4 + \lambda_{4; R} + 12[\lambda_2 - 2\lambda_1^2 + \lambda_{2; R}] \bar{\mu}_{2; \mu_1} \\ & + 24[\bar{\mu}_{11; \mu_1 \mu_2} - \bar{\mu}_{3; \mu_1}] \lambda_1 - 4\bar{\mu}_{11; \mu_1 \mu_3} \\ & + 12\bar{\mu}_{21; \mu_1 \mu_2} - 6\bar{\mu}_{4; \mu_1} - 3\bar{\mu}_{2; \mu_2}. \end{aligned}$$

Formulas for Thiele seminvariants of ungrouped data in terms of expectations may be obtained from (2.19) through (2.21).

Adjustment formulas for Thiele seminvariants of a distribution of continuous variates are given by Langdon and Ore [23; p. 231] and Craig [25; p. 57]. If we denote the t th Thiele seminvariant of a distribution of continuous variates by L_t , then

$$(2.22) \quad v_{1; L_t} = L_t + L_{t; R},$$

where

$$(2.23) \quad L_{2t+1; R} = 0, \quad L_{2t; R} = \frac{(-1)^{t-1} B_t m^{2t}}{2t}, \quad t = 1, 2, \dots$$

Formulas (2.19) through (2.21) may be used for continuous variates by changing the moment symbols and setting terms of the form $\bar{\mu}_{s_1 s_2 \dots s_r; \mu_1 \mu_2 \dots \mu_r}$ equal to zero.

6. Adjustment formulas applied to a numerical problem. We consider the arbitrary distribution given in Table III.

TABLE III
An Arbitrary Distribution of Discrete Variates

v	f	v	f	v	f	F
1	2	4	30	7	1	$2 + 30 + 1 = 33$
2	8	5	4	8	1	$8 + 4 + 1 = 13$
3	10	6	3	9	1	$10 + 3 + 1 = 14$

The three grouped distributions, when the variates are grouped in "groupings of three," appear in Table IV.

TABLE IV

Distributions Derived from Data of Table III by Making the Three Possible Groupings of Three

(1)		(2)		(3)	
Class	f	Class	f	Class	f
1-3	20	0-2	10	-1 to 1	2
4-6	37	3-5	44	2-4	48
7-9	3	6-8	5	5-7	8
10-12	0	9-11	1	8-10	2

Using the fixed point 4, moment-functions are computed for the distribution of Table III and for each of the distributions of Table IV. These quantities along with the average of each moment function appear in Table V.

TABLE V

Moment-Functions of the Distributions of Table III and Table IV. Averages of Moment-Functions of Distributions of Table IV

Dist.	μ_1	μ_2	μ_3	μ_4	$\bar{\mu}_2 = \lambda_2$	$\bar{\mu}_3 = \lambda_3$	$\bar{\mu}_4$	λ_4
(1)	$\frac{9}{60}$	$\frac{165}{60}$	$\frac{69}{60}$	$\frac{1125}{60}$	$\frac{9819}{(60)^2}$	$\frac{-17442}{(60)^3}$	$\frac{238,849,317}{(60)^4}$	$\frac{-50,388,966}{(60)^4}$
(2)	$\frac{-9}{60}$	$\frac{171}{60}$	$\frac{81}{60}$	$\frac{2511}{60}$	$\frac{10179}{(60)^2}$	$\frac{567162}{(60)^3}$	$\frac{557,840,277}{(60)^4}$	$\frac{247,004,154}{(60)^4}$
(3)	$\frac{-30}{60}$	$\frac{162}{60}$	$\frac{138}{60}$	$\frac{1938}{60}$	$\frac{8820}{(60)^2}$	$\frac{1317600}{(60)^3}$	$\frac{528,282,000}{(60)^4}$	$\frac{294,904,800}{(60)^4}$
Ave.	$\frac{-10}{60}$	$\frac{166}{60}$	$\frac{96}{60}$	$\frac{1858}{60}$	$\frac{9606}{(60)^2}$	$\frac{622440}{(60)^3}$	$\frac{441,657,198}{(60)^4}$	$\frac{163,839,996}{(60)^4}$
Orig. Dist.	$\frac{-10}{60}$	$\frac{126}{60}$	$\frac{116}{60}$	$\frac{1314}{60}$	$\frac{7460}{(60)^2}$	$\frac{642400}{(60)^3}$	$\frac{305,034,000}{(60)^4}$	$\frac{138,079,200}{(60)^4}$

Table VI gives the expected values of the moment-functions as obtained by substituting from Table V into the formulas of sections two, three, and five. Also the expected values, computed from the usual formulas, are given and the errors which would be made, if the usual formulas were used, are indicated.

TABLE VI
Expected Values of Moment-Functions Computed by Formulas

Expectations by	$\mu_{1:\mu_1}$	$\mu_{1:\mu_2}$	$\mu_{1:\mu_3}$	$\mu_{1:\mu_4}$	$\mu_{1:\bar{\mu}_2} = \mu_{1:\lambda_2}$	$\mu_{1:\bar{\mu}_3} = \mu_{1:\lambda_3}$	$\mu_{1:\bar{\mu}_4}$	$\mu_{1:\lambda_4}$
New Formulas	$\frac{-10}{60}$	$\frac{166}{60}$	$\frac{96}{60}$	$\frac{1858}{60}$	$\frac{9606}{(60)^2}$	$\frac{622440}{(60)^3}$	$\frac{441,657,198}{(60)^4}$	$\frac{163,839,996}{(60)^4}$
Usual Formulas	$\frac{-10}{60}$	$\frac{166}{60}$	$\frac{96}{60}$	$\frac{1858}{60}$	$\frac{9860}{(60)^2}$	$\frac{642400}{(60)^3}$	$\frac{416,778,000}{(60)^4}$	$\frac{133,795,200}{(60)^4}$
Error	—	—	—	—	$\frac{254}{(60)^2}$	$\frac{19960}{(60)^3}$	$\frac{-24,879,198}{(60)^4}$	$\frac{-30,060,796}{(60)^4}$

7. Evaluation of $\bar{\mu}_{2:\mu_1}$. It appears at first that it is necessary to form the " k groupings of k " in order to evaluate the term $\bar{\mu}_{2:\mu_1}$ which enters the precise formula for the expected value of the variance. That was the procedure followed by Carver [26; p. 161]. However, it is possible to evaluate $\bar{\mu}_{2:\mu_1}$ from the ungrouped data without forming a single grouped-distribution.

By definition,

$$\bar{\mu}_{2:\mu_1} = \frac{1}{k} \sum_{r=1}^k [r\mu_1 - \mu_1]^2,$$

where $r\mu_1$ is the mean of the r th grouped-distribution and μ_1 is the mean of the ungrouped distribution. We wish to study the terms $r\mu_1$ and μ_1 . Consider a set of variates x_i , ($i = 1, 2, \dots, s$), with corresponding frequencies f_i , ($i = 1, 2, \dots, s$). The x 's are subject to the condition, $x_i - x_{i-1} = 1$, and consequently some of the f 's may be zero. The mean of this distribution is $\frac{\sum xf}{\sum f}$.

We define

$$F_i = f_i + f_{k+i} + f_{2k+i} + \dots, \quad (i = 1, 2, \dots, k).$$

Then, if a grouped-distribution is formed with x_i in the i th ($i = 1, 2, \dots, k$) position of a class, the mean of this grouped-distribution is

$$\frac{\sum xf + \sum_{j=1}^k F_j e_{i+j-1}}{\sum f}$$

where $e_{i-1} = e_k$ if $e_i = 1$ and $e_{i+1} = e_1$ if $e_i = e_k$. Similarly if a grouped-distribution is formed with x_i in the $(i+1)$ st position of a class, the mean is

$$\frac{\sum xf + \sum_{j=1}^k F_j e_{i+j}}{\sum f}.$$

Thus, it is evident that, given the expression for the mean of any grouped-distribution in which x_i is in the i th position of a class, we may form the expression for the mean of the grouped-distribution in which x_i is in the $(i + 1)$ st position of a class by a cyclic permutation of the e_i 's of the given expression.

Therefore, it follows that if we call ${}_r\mu_1$ the mean of the grouped-distribution in which x_i is in the r th ($r = 1, 2, \dots, k$) position of a class, then

$${}_r\mu_1 - \mu_1 = \frac{\sum_{j=1}^k F_j e_{r+j-1}}{\sum f}, \quad (r = 1, 2, \dots, k).$$

If we define

$$N = \sum f \quad \text{and} \quad \phi_r = \sum_{j=1}^k F_j e_{r+j-1}$$

then,

$$\bar{\mu}_{2;\mu_1} = \frac{1}{kN^2} \sum_{r=1}^k \phi_r^2.$$

Thus, it is evident that $\bar{\mu}_{2;\mu_1}$ is a function of the frequencies of the variates and of the e_i 's. The fact that the values of the variates do not enter $\bar{\mu}_{2;\mu_1}$ permits one to quickly calculate its value.

Consider $\bar{\mu}_{2;\mu_1}$ for the distribution of Table III. We find

$$\phi_1 = 33e_1 + 13e_2 + 14e_3.$$

Then, by successive cyclic permutations of the e_i 's,

$$\phi_2 = 33e_2 + 13e_3 + 14e_1,$$

$$\phi_3 = 33e_3 + 13e_1 + 14e_2.$$

Substituting the values $e_1 = 1, e_2 = 0, e_3 = -1$ we have $\phi_1 = 19, \phi_2 = 1$ and $\phi_3 = -20$. Therefore,

$$\bar{\mu}_{2;\mu_1} = \frac{254}{(60)^2}$$

which is identical with the value which was found when Table V was used.

It follows from the preceding development that

$$\bar{\mu}_{t;\mu_1} = \frac{1}{kN^t} \sum_{r=1}^k \phi_r^t$$

and if $F_1 = F_2 = \dots = F_k$ then $\bar{\mu}_{t;\mu_1}$ is zero.

8. Conclusion. The results of this paper include:

1. The derivation of general and specific formulas for the expected values of population moment-functions.

2. The derivation of generalized sampling formulas under the condition that samples of n are formed by selecting one variate from each population.

3. Methods for the transformation of generalized sampling formulas into the corresponding infinite and finite sampling formulas.

4. A method for the transformation of infinite sampling formulas into the corresponding finite sampling formulas.

5. A demonstration of the fact that adjustment formulas for moment-function of grouped data involve central moments of a rectangular distribution.

6. A general formula for the expected value of the t th moment of grouped data.

7. New adjustment formulas for central moments of grouped data.

8. New adjustment formulas for Thiele seminvariants of grouped data.

9. A method for the evaluation of the term $\bar{\mu}_{2;\mu_1}$ which appears in the precise adjustment formula for the variance.

Many thanks are due Prof. P. S. Dwyer, to whom the writer is greatly indebted for advice and encouragement.

REFERENCES

- [1] "STUDENT", "The Probable Error of the Mean," *Biometrika*, Vol. 6 (1908), pp. 1-25.
- [2] L. ISSERLIS, "On the Value of a Mean as Calculated from a Sample," *Jour. of Roy. Stat. Soc.*, Vol. 81 (1918), pp. 75-81.
- [3] A. TCHOUPROFF, "On the Mathematical Expectation of the Moments of Frequency Distributions," Part I, *Biometrika*, Vol. 12 (1918), pp. 140-169, 185-210; Part II, *Biometrika*, Vol. 13 (1920-21), pp. 283-295.
- [4] A. TCHOUPROFF, "On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations," *Metron*, Vol. 2 (1923), pp. 461-493; 646-683.
- [5] A. CHURCH, "On the Moments of the Distributions of Squared Deviations for Samples of N Drawn from an Indefinitely Large Population," *Biometrika*, Vol. 17 (1925), pp. 79-83.
- [6] J. NEYMAN, "Contributions to the Theory of Small Samples Drawn from a Finite Population," *Biometrika*, Vol. 17 (1925), pp. 472-479.
- [7] A. CHURCH, "On the Means and Squared Deviations of Small Samples from any Population," *Biometrika*, Vol. 18 (1926), pp. 321-394.
- [8] K. PEARSON, "Another Historical Note on the Problem of Small Samples," *Biometrika* Vol. 19 (1927), pp. 207-210.
- [9] C. C. CRAIG, "An Application of Thiele's Seminvariants to the Sampling Problem," *Metron*, Vol. 7 (1928-29), pp. 3-74.
- [10] R. A. FISHER, "Moments and Product Moments of Sampling Distributions," *Proc. London Math. Soc.*, Vol. 2 (30) (1929), pp. 199-238.
- [11] J. PEPPER, "Studies in the Theory of Sampling," *Biometrika*, Vol. 11 (1929), pp. 231-258.
- [12] H. C. CARVER, "Fundamentals of Sampling," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 101-121, 260-274.
- [13] N. ST. GEORGESQUE, "Further Contributions to the Sampling Problem," *Biometrika*, Vol. 24 (1932), pp. 65-107.
- [14] G. M. BROWN, "On Sampling from Compound Populations," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 288-342.
- [15] J. NEYMAN, "On the Two Different Aspects of the Representative Method," *Jour. Roy. Stat. Soc.*, Vol. 97 (1934), pp. 558-625.

- [16] A. L. BOWLEY, "Measurement of the Precision Attained in Sampling," *Bull. International Stat. Inst.*, Vol. 22.
- [17] H. FELDMAN, "Mathematical Expectations of Product Moments of Samples Drawn from a Set of Infinite Populations," *Annals of Math. Stat.*, Vol. 6 (1935), pp. 30-52.
- [18] H. C. CARVER, "Sheppard's Corrections for a Discrete Variable," *Annals of Math. Stat.*, Vol. 76 (1936), pp. 154-163.
- [19] P. S. DWYER, "Moments of Any Rational, Integral Isobaric Sample Moment Function," *Annals of Math. Stat.*, Vol. 8 (1937), pp. 21-65.
- [20] P. S. DWYER, "Combined Expansions of Products of Symmetric Power Sums and of Sums of Symmetric Power Products with Application to Sampling," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 1-84.
- [21] A. T. CRAIG, "On the Mathematics of the Representative Method of Sampling," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 26-34.
- [22] A. FISHER, *The Mathematical Theory of Probabilities*, New York, The Macmillan Company. 1928.
- [23] W. H. LANGDON, and ORE, O., "Seminvariants and Sheppard's Corrections," *Annals of Math.*, Vol. 31 (1930), pp. 230-232.
- [24] J. R. ABERNETHY, "On the Elimination of Systematic Errors Due to Grouping," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 263-277.
- [25] C. C. CRAIG, "Sheppard's Corrections for a Discrete Variable," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 55-61.
- [26] H. C. CARVER, "The Fundamental Nature and Proof of Sheppard's Adjustments," *Annals of Math. Stat.*, Vol. 9 (1936), pp. 154-163.
- [27] M. SASULY, *Trend Analysis of Statistics*, Washington, The Brookings Institution (1934).

THE UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN

THE ANALYSIS OF VARIANCE WHEN EXPERIMENTAL ERRORS FOLLOW THE POISSON OR BINOMIAL LAWS

By W. G. COCHRAN

1. Introduction. The use of transformations has recently been discussed by several writers [1], [2], [3], [4], in applying the analysis of variance to experimental data where there is reason to suspect that the experimental errors are not normally distributed. Two types of transformations appear to be coming into fairly common use: \sqrt{x} and $\sin^{-1} \sqrt{x}$. The former is considered appropriate where the data are small integers whose experimental errors follow the Poisson law, while the latter applies to fractions or percentages derived from the ratio of two small integers, where the experimental errors follow the binomial frequency distribution. In each case the object of the transformation is to put the data on a scale in which the experimental variance is approximately the same on all plots, so that all plots may be used in estimating the standard error of any treatment comparison. The extent to which these transformations are likely to succeed in so doing has been examined by Bartlett [2]. The object of the present paper is to discuss the theoretical basis for these transformations in more detail, and in particular to examine their relation to a more exact analysis.

2. Experimental variation of the Poisson type. The first step in an exact statistical analysis of the results of any field experiment, is to specify in mathematical terms (1) how the expected values on each plot are obtained in terms of unknown parameters representing the treatment and block (or row and column) effects (2) how the observed values on the plots vary about the expected values. In this section, the variation is assumed to follow the Poisson law.

The specification of the expected values requires some consideration. In the standard theory of the analysis of variance, treatment and block (or row and column) effects are assumed to be additive. In the case of a Latin square, for example, the expected yield m_i of the i th plot, which receives the t th treatment and occurs in the r th row and the c th column is written

$$(1) \quad m_i = G + T_t + R_r + C_c$$

where G is a parameter representing the average level of yield in the experiment, and T_t , R_r and C_c represent the respective effects of the treatment, row and column to which the plot corresponds. Since the T , R and C constants are required only to measure differences between different treatments, rows and columns, we may put

$$(2) \quad \sum_i T_t = \sum_r R_r = \sum_c C_c = 0.$$

If the experimental errors are normally and independently distributed with equal variance, this specification leads to very simple equations of estimation for the unknown parameters, the maximum likelihood estimate of T_i , for example, being the difference between the mean yield of all plots receiving that treatment and the general mean. In addition to its simplicity, this type of prediction formula is fairly suitable for general use, because it gives a good approximation to most types of law which might be envisaged, provided that row and column differences are small in relation to the mean yield. However, in considering an exact analysis with Poisson variation, the prediction formula is assumed chosen, without reference to computational simplicity, as being the most suitable to describe the combined actions of treatment and soil effects.

The probability of obtaining a given set of plot yields x_i with expectations m_i may be written

$$\prod_i \frac{e^{-m_i} m_i^{x_i}}{x_i!}.$$

Thus L , the logarithm of the likelihood, is given by

$$(3) \quad L = \sum_i (x_i \log m_i - m_i) - \sum_i \log x_i!$$

Hence the maximum likelihood equation of estimation for any parameter θ assumes the form

$$(4) \quad \sum \frac{(x_i - m_i)}{m_i} \frac{\partial m_i}{\partial \theta} = 0$$

where the summation extends over all plots whose expectations involve θ . The function $\frac{\partial m_i}{\partial \theta}$ will usually involve a number of parameters. Since the specification of row, column and treatment effects in a 6 x 6 Latin square requires 16 independent parameters, the solution of these equations may be expected to be laborious, though it may be shortened by the intelligent use of iterative methods. The problem of obtaining exact tests of significance is also difficult. The method of maximum likelihood provides estimates of the variances and covariances of the treatment constants, which under certain conditions can be assumed to be normally distributed if there is sufficient replication, but this can hardly be considered an exact "small sample" solution.

These remarks show that the exact solution is somewhat too complicated for frequent use. The difficulty arises principally because the typical equation of estimation consists of a *weighted* sum of the deviations of the observed from the expected values, the weights being $\frac{1}{m_i} \frac{\partial m_i}{\partial \theta}$. The factor $\frac{1}{m_i}$ was introduced into the weight by the Poisson variation of the experimental errors, and must be retained in any theory which claims to apply to Poisson variation. It is, however, worth considering whether some simplification cannot be introduced into

the equations by assuming some particular form for the prediction formula. This line of approach seems promising when one considers the simplification introduced into the "normal theory" case by assuming the prediction formula to be linear.

For Poisson variation, the linear law does not appear to be particularly suitable, since it may give negative expectations on some plots (as happens in the numerical example considered in the next section). Further, while $\frac{\partial m_i}{\partial \theta}$ becomes a constant, the factor $\frac{1}{m_i}$ remains in the weight.

The entire weight can be made constant by assuming a linear prediction formula in the square roots and transforming the data to square roots. For a Latin square, this prediction formula is written

$$(5) \quad \sqrt{m_i} = \alpha_i = G + T_t + R_r + C_c,$$

where

$$(6) \quad \sum_t T_t = \sum_r R_r = \sum_c C_c = 0.$$

To find the maximum value of (3) subject to the restrictions (6), we may use the method of undetermined multipliers, maximizing

$$(7) \quad L + \lambda(\sum_t T_t) + \mu(\sum_r R_r) + \nu(\sum_c C_c).$$

The equation of estimation for a typical treatment constant T_t becomes

$$(8) \quad \sum \left(\frac{x_i - m_i}{m_i} \right) \frac{dm_i}{d\alpha_i} \frac{\partial \alpha_i}{\partial T_t} + \lambda = 0, \quad \text{i.e.,} \quad \sum \frac{2(x_i - m_i)}{\sqrt{m_i}} + \lambda = 0,$$

the summation being extended over all plots receiving the treatment. If $a_i = \sqrt{x_i}$, then by Taylor's theorem

$$(9) \quad x_i - m_i = (a_i - \alpha_i) \frac{dm_i}{d\alpha_i} + \frac{1}{2!} (a_i - \alpha_i)^2 \frac{d^2 m_i}{d\alpha_i^2} + \dots$$

If m_i is reasonably large, only the first term on the right-hand side need be retained. When m_i is small, we may use, instead of the exact square root, a quantity a'_i defined so that

$$(10) \quad x_i - m_i = (a'_i - \alpha_i) \frac{dm_i}{d\alpha_i} = 2\sqrt{m_i}(a'_i - \alpha_i).$$

Thus if the analysis is performed on the quantities a'_i instead of on the original data, equation (8) becomes

$$(11) \quad \sum_{\tau_i} 4(a'_i - \alpha_i) + \lambda = 0.$$

On substituting the expectations for α_i from (5), and using (6), we obtain

$$(12) \quad \sum_i 4(a'_i - G - T_i) + \lambda = 0.$$

The corresponding equation for G is

$$(13) \quad \sum_i 4(a'_i - G) = 0,$$

so that G is the general mean of the quantities a' . By adding equations (12) over all treatments, and comparing the total with (13), we find $\lambda = 0$. Hence T_i is the difference between the mean yield of a' over all plots receiving T_i and the general mean of a' . In this scale the simplicity of the "normal theory" equations has apparently been recovered. Actually, the quantities a' are not known exactly, since

$$(14) \quad a' = \alpha + \frac{(x - m)}{2\sqrt{m}} = \frac{1}{2} \left(\alpha + \frac{x}{\alpha} \right)$$

where α is the expected value of \sqrt{x} . However, this process provides a means of successively approximating the maximum likelihood solution, by choosing first approximations to the quantities α , constructing the a' 's, solving for the unknown constants and hence obtaining second approximations to the expected values. The close relation of a' to \sqrt{x} is seen by remembering one of the common rules for finding square roots. This consists in guessing an approximate root (α), dividing x by the approximate root, and taking the mean of the approximate root (α) and the resulting quotient (x/α).

The suitability of the linear prediction formula in square roots must be considered in any example in which the above analysis is being employed. The law is intermediate in its effects between the linear law and the product law in the original data. My experience is that it is fairly satisfactory for general use, (cf. [2], p. 72). An exception may occur when it is desired to test the interaction between two treatments, both of which produce large effects. In this case the definition chosen for absence of interaction may not coincide at all closely with the definition implied in using the linear law in square roots. An example of this case was given in a previous paper [1].

In this connection it should be noted that an approximate "goodness of fit" test may be obtained of the validity of the assumptions made. Since the quantities a'_i enter into the equations of estimation with weight 4, the quantity $4 \sum_i (a'_i - \alpha_i)^2$ is distributed approximately as χ^2 with the number of degrees of freedom in the error term of the analysis of variance. Some idea of the closeness of the approximation may be gathered by considering the simplest case in which only the mean yield is being estimated. In this case the observed values x are assumed to be drawn from the same Poisson distribution, and the sufficient statistic for the mean G is known to be $\Sigma(x_i)/n$. Since, however, the

prediction formula is here the same in square roots as in the original scale, and since the maximum likelihood solution is invariant to change of scale, the mean value α of a' must be *exactly* $\sqrt{\Sigma(x)/n}$, as the reader may verify by working any particular example. Thus $\Sigma 4(a' - \alpha)^2$ is found to be $\Sigma(x - \bar{x})^2/\bar{x}$, the usual χ^2 test for examining whether a set of values x may reasonably be assumed to come from the same Poisson distribution. By working out the exact distribution of $\Sigma(x_i - \bar{x})^2/\bar{x}$ in a number of cases [5], I previously expressed the opinion that this quantity followed the χ^2 distribution sufficiently closely for most practical uses, even for values of the mean as low as 2. This opinion has since been substantiated by Sukhatme, [6] who sampled this distribution for $m = 1, 2, 3, 4$, and 5.

A high value of χ^2 means either that the prediction formula is not satisfactory or that the experimental errors are higher than the Poisson distribution indicates, or that both causes are operating. These effects can sometimes be separated by examining whether the observed yields deviate from the expected yields in a systematic or a random manner. If the deviation is systematic, the prediction formula is probably unsatisfactory.

The type of approach used above resembles in many features the "exact" analysis for the probit transformation [7]. The principal difference is that in the case of probits the transformation is made to suit the *a priori* prediction formula, which postulates that the probits are a linear function of the dosage, or of the log (dosage). Thus with probits the equations of estimation still involve weights in the transformed scale. These do not seriously complicate the analysis, since only two parameters require to be estimated for a given poison. With, however, the much greater number of parameters usually involved in specifying the results of a field experiment, the attractiveness of a solution which does not involve weighting is greatly increased.

3. Numerical example of the square root transformation. A 5×5 Latin square experiment on the effects of different soil fumigants in controlling wireworms was selected as an example. The average number of wireworms per plot (total of four soil samples) was just under five. Previous studies [8], [9] have indicated that with small numbers per sample, the distribution of numbers of wireworms tends to follow the Poisson law.

The plan and yields are shown in Table I. The first two figures under the treatment symbols are the numbers of wireworms and their square roots respectively, the latter being regarded as first approximations to the values a' . Two of the plots receiving treatment K gave no wireworms. Since these plots are likely to be changed most in the transition from square roots to a' , better approximations were estimated for them before proceeding with the calculations. The best simple approximations appeared to be obtained from the square roots of the means in the original units. For the plot in the second row and second column, the square roots of the row, column and treatment means in the original

TABLE I
Plan and number of wireworms per plot

<i>P</i>	<i>O</i>	<i>N</i>	<i>K</i>	<i>M</i>	Mean
3 ¹	2	5	1	4	
1.73 ²	1.41	2.24	1.00	2.00	1.676 ²
1.76 ³	1.45	2.25	1.11	2.00	1.714 ³
1.77 ⁴	1.46	2.25	1.10	2.00	1.716 ⁴
<i>M</i>	<i>K</i>	<i>O</i>	<i>N</i>	<i>P</i>	
6	0	6	4	4	
2.45	(0.39)	2.45	2.00	2.00	1.858
2.45	0.32	2.50	2.02	2.02	1.862
2.46	0.32	2.49	2.02	2.02	1.862
<i>O</i>	<i>M</i>	<i>K</i>	<i>P</i>	<i>N</i>	
4	9	1	6	5	
2.00	3.00	1.00	2.45	2.24	2.138
2.10	3.09	1.00	2.47	2.25	2.182
2.13	3.08	1.00	2.46	2.25	2.184
<i>N</i>	<i>P</i>	<i>M</i>	<i>O</i>	<i>K</i>	
17	8	8	9	0	
4.12	2.83	2.83	3.00	(0.79)	2.714
4.18	2.84	2.83	3.00	0.77	2.724
4.17	2.84	2.83	3.00	0.77	2.722
<i>K</i>	<i>N</i>	<i>P</i>	<i>M</i>	<i>O</i>	
4	4	2	4	8	
2.00	2.00	1.41	2.00	2.83	2.048
2.14	2.02	1.49	2.04	2.92	2.122
2.10	2.03	1.50	2.05	2.90	2.116
Mean					
2.460 ²	1.926	1.986	2.090	1.972	2.087 ²
2.526 ³	1.944	2.014	2.128	1.992	2.121 ³
2.526 ⁴	1.946	2.014	2.126	1.988	
Treatment Means					
<i>K</i>	<i>P</i>	<i>O</i>	<i>M</i>	<i>N</i>	
1.036 ²	2.084	2.338	2.456	2.520	
1.068 ³	2.116	2.394	2.482	2.544	
1.058 ⁴	2.118	2.396	2.484	2.544	

¹Original numbers. ²Square roots. ³Second approximations. ⁴Third approximations.

units are respectively 2.000, 2.145 and 1.095, and the square root of the general mean is 2.227. Hence

$$a' = \frac{1}{2}[2.000 + 2.145 + 1.095 - 2(2.227)] = 0.39.$$

The other zero value was similarly found to give $a' = 0.79$. The corresponding estimates from the means of the square roots were considerably too low, since the a' values tend to be higher than the square roots. The use of "missing plot" technique gave very poor approximations, because it ignores the fact that the plots in question had zero yields.

With the estimated values inserted, the row, column, and treatment means of the square roots are as shown in Table I. A second approximation to a' was calculated for each plot. For the plot in the first row and the first column, the expected yield is

$$\alpha = 1.676 + 2.460 + 2.084 - 2(2.087) = 2.046.$$

Hence $a' = \frac{1}{2}(2.046 + 3/2.046) = 1.76$. These values constitute the third set of figures in Table I. Theoretically, it is advisable to readjust the row, column, and treatment means after each new value of a' has been obtained, in order to secure rapid convergence. This is rather laborious in practice, and a complete set of new plot values was obtained before readjusting the means. The third approximations obtained by this method are shown in the fourth lines in Table I and are correct to two decimal places.

It is noteworthy how closely the square roots agree with the third approximations on all plots except those which originally gave zero yields. The differences between the second and third approximations are trivial.

The next step is to make a χ^2 test by means of the quantity $4\Sigma(a' - \alpha)^2$. From the manner in which the values α are constructed from the a'' 's, it follows that $\Sigma(a' - \alpha)^2$ is simply the error sum of squares in the conventional analysis of variance of the values a' . The analysis of variance of the third approximations is shown in Table II.

TABLE II
Analysis of variance of adjusted square roots

	Degrees of freedom	Sum of squares	Mean square
Rows	4	2.9815	
Columns	4	1.1190	
Treatments	4	7.5815	1.8954
Error	12	4.5970	0.3831

The value of χ^2 is $4 \times 4.597 = 18.39$, with 12 degrees of freedom, which is just about the 10 percent level. If the hypothesis is regarded as disproved only when χ^2 exceeds the 5 percent level, the treatment means may be tested by regarding them as approximately normally distributed with variance

$1/5 \times 0.25 = 0.05$. It is, however, more prudent to use the actual error mean square as an estimate of the experimental error variance, performing the usual tests associated with the analysis of variance. This may be justified on the grounds that the calculations have produced a set of plot values a' of equal weight. On this basis the standard error of a treatment mean is $\sqrt{0.3831/5} = 0.2768$. Treatment K reduced the number of wireworms significantly below all other treatments, but there is no indication of any difference between the other treatments. The treatment means may be reconverted to the original units by squaring.

4. Experimental variation of the binomial type. In this case the yields are obtained by examining a constant number n units per plot and noting those which possess a certain attribute (e.g., plants which are diseased). Experimental variation is presumed to arise solely from the binomial variation of the observed fraction p possessing the attribute about the expected fraction P , which is specified in terms of unknown parameters representing the treatment and soil effects.

If r_i is the number possessing the attribute on a typical plot, so that $p_i = r_i/n$ the likelihood function takes the form

$$\prod_i \frac{n!}{r_i!(n-r_i)!} P_i^{r_i} Q_i^{n-r_i}.$$

Hence the terms in the logarithm which involve the unknown parameters are given by

$$(15) \quad L = \sum_i \{r_i \log P_i + (n - r_i) \log Q_i\}.$$

The equation of estimation for a typical constant θ is

$$(16) \quad \sum \frac{n}{P_i Q_i} (p_i - P_i) \frac{\partial P_i}{\partial \theta} = 0$$

where the summation is over all plots whose expectations involve θ .

As in the Poisson case, an exact solution is laborious because of the weights $\frac{n}{P_i Q_i} \cdot \frac{\partial P_i}{\partial \theta}$. The unequal weighting may be removed by transforming to the variate $\alpha_i = \sin^{-1} \sqrt{P_i}$, and assuming that the prediction formula is linear in the transformed scale. For a Latin square the prediction formula is assumed to be

$$(17) \quad \alpha_i = G + T_t + R_r + C_c$$

where the i th plot receives treatment t and lies in the r th row and c th column. Further

$$(18) \quad \sum_i T_t = \sum_r R_r = \sum_c C_c = 0.$$

Since $P_i = \sin^2 \alpha_i$, $\frac{dP_i}{d\alpha_i} = 2\sqrt{P_i Q_i}$. A set of variates a'_i is defined so that on each plot

$$(19) \quad p_i - P_i = (a'_i - \alpha_i) \frac{dP_i}{d\alpha_i} = 2\sqrt{P_i Q_i} (a'_i - \alpha_i).$$

With these substitutions, the equation of estimation for T_i , for instance, becomes

$$(20) \quad \sum_{T_i} 4n(a'_i - \alpha_i) + \lambda = 0$$

where, as before, λ is an undetermined multiplier. The remainder of the solution proceeds exactly as in the Poisson case, T_i being found to be the difference between the mean value of a'_i over all plots receiving this treatment and the general mean of a'_i . A χ^2 test may be made with $\sum_i 4n(a'_i - \alpha_i)^2$.

From (19)

$$(21) \quad a'_i = \alpha_i + \frac{1}{2\sqrt{P_i Q_i}} (p_i - P_i) = \alpha_i + \frac{1}{2\sqrt{P_i Q_i}} (Q_i - q_i)$$

$$(22) \quad = \alpha_i + \frac{1}{2} \cot \alpha_i - q_i \operatorname{cosec} (2\alpha_i)$$

where q_i is the observed fraction which does not possess the attribute. The calculation of approximations to a'_i thus involves finding a predicted value α_i from the treatment and block (or row and column) means, and using equation (22). Tables [10] of the values of $\sin^{-1} \sqrt{P_i}$, $\alpha_i + \frac{1}{2} \cot \alpha_i$, and $\operatorname{cosec} (2\alpha_i)$ have been prepared to facilitate the computations. It should be noted that these tables are in degrees, whereas the above equations assume that α_i is measured in radians. In degrees, equation (20) above becomes

$$(23) \quad \sum_{T_i} \frac{\pi^2 n}{8100} (a'_i - \alpha_i) = 0$$

while

$$(24) \quad a'_i = \alpha_i + \frac{180}{\pi} \left\{ \frac{1}{2} \cot \alpha_i - q_i \operatorname{cosec} (2\alpha_i) \right\}.$$

As in the Poisson case, the appropriateness of the linearly additive law in equivalent angles depends on the way in which treatment and soil effects operate. As Bliss has shown [11], the effect of the transformation is to flatten out the cumulative normal frequency distribution, extending the range over which it can be approximated by a straight line.

5. Numerical example of the angular transformation. The data were selected from a randomized blocks experiment by Carruth [12] on the control by mechanical and insecticidal methods of damage due to corn ear worm larvae.

The control and the six types of mechanical protection were chosen for analysis, the "yields" being the percentages of ears unfit for sale. The numbers of ears varied somewhat from plot to plot, the average being 36.5, but the variations were fairly small and appeared to be random. It was considered that variations in the weight ($4n$) could be ignored in solving the equations of estimation.

TABLE III
Percentages of unfit ears of corn

Treatments	Blocks						Means
	I	II	III	IV	V	VI	
1	42.4 ¹	34.3	24.1	39.5	55.5	49.1	
	40.6 ²	35.8	29.4	38.9	48.2	44.5	39.57 ²
	40.7 ³	36.0	29.4	38.9	48.6	44.6	39.70 ³
2	23.5	15.1	11.8	9.4	31.7	15.9	
	29.0	22.9	20.1	17.9	34.3	23.5	24.62
	29.1	23.1	20.3	18.2	34.3	23.5	24.75
3	33.3	33.3	5.0	26.3	30.2	28.6	
	35.2	35.2	12.9	30.9	33.3	32.3	29.97
	35.5	35.3	14.5	31.0	33.4	32.4	30.35
4	11.4	13.5	2.5	16.6	39.4	11.1	
	19.7	21.6	9.1	24.0	38.9	19.5	22.13
	19.8	21.7	10.0	24.4	39.9	19.6	22.57
5	14.3	29.0	10.8	21.9	30.8	15.0	
	22.2	32.6	19.2	27.9	33.7	22.8	26.40
	22.6	32.7	19.2	28.0	33.7	22.9	26.52
6	8.5	21.9	6.2	16.0	13.5	15.4	
	17.0	27.9	14.4	23.6	21.6	23.1	21.27
	17.4	28.2	14.5	24.0	22.1	23.2	21.57
7	16.6	19.3	16.6	2.1	11.1	11.1	
	24.0	26.1	24.0	8.3	19.5	19.5	20.23
	24.3	26.2	28.8	10.9	20.1	19.5	21.63
Means	26.81 ²	28.87	18.44	24.50	32.79	26.46	26.31

¹ Percentage. ² Equivalent angle. ³ Second approximation.

The percentages of unfit ears, the equivalent angles and the second approximations to a' are shown in descending order in Table III. The percentages on

individual plots vary from 2.1 to 55.5. The second approximations were calculated from the block and treatment means of the angles. For the control plot (treatment 1) in block I, for example, the expected value is

$$39.57 + 26.81 - 26.31 = 40.07.$$

Since Fisher and Yates's tables of $\alpha + \frac{1}{2} \cot \alpha$ and $\operatorname{cosec} (2\alpha)$ are given for values of α from 45° to 90° , we take the complement of the expected value, which is 49.93. Interpolating mentally from the table, we find

$$\alpha + \frac{1}{2} \cot \alpha = 74.0, \operatorname{cosec} (2\alpha) = 58.3.$$

Thus the second approximation to the complement of the angle is

$$74.0 - 0.424 \times 58.3 = 49.3.$$

Hence the second approximation to a' is 40.7, which agrees very closely with the equivalent angle.

On the majority of the plots, the second approximation differs by only a trivial amount from the equivalent angle. The plots with the three lowest percentages (2.1, 2.5, and 5.0) have increased somewhat more, and also one or two other plots where the angles deviated considerably from the expected values. A third set of approximations was not considered necessary.

The analysis of variance of the second approximations is given in Table IV.

TABLE IV

	Degrees of freedom	Sum of squares	Mean squares
Blocks	5	709.79	
Treatments	6	1,531.56	255.26
Error	30	982.67	32.76

Taking n as 36.5, the expected value of the error mean square is $820.7/36.5 = 22.48$. Thus $\chi^2 = 982.67/22.48 = 43.71$, with 30 degrees of freedom, which is almost exactly at the 5 percent level. This, together with the appreciable amount of the variance removed by blocks, indicates that the experimental error probably contains some element other than binomial variation. As in the preceding case, it would be wise to make the usual analysis of variance tests with the actual error mean square.

6. Discussion. It must be emphasized that the solutions given above apply to the case where the whole of the experimental error variation is of the Poisson or binomial type. The methods are therefore likely to be useful in practice only where the experimental conditions have been carefully controlled, or where the data are derived from such small numbers that the Poisson or binomial variation is much larger than any extraneous variation. The χ^2 test is helpful in deciding

whether this assumption is justified. Further, the examples worked above indicate that the transformed values form very good approximations on most plots. It will often be sufficient to adjust only those plots which give zero or very small values in the Poisson case, or zero or 100 percent values in the binomial case. In this connection the method of adjustment given above may perhaps be considered as an improvement on the empirical rule given by Bartlett [13] of counting n out of n as $(n - 1/4)$ out of n .

Where extraneous variation becomes important, as is probably the normal case with data derived from field experiments, there seem to be no theoretical grounds for using the adjusted values. If we were prepared to describe accurately the nature of the variation other than that of the Poisson or binomial type, a new set of maximum likelihood equations could be developed. These would, however, lead to a different type of adjustment.

The justification for the use of transformations has no direct relation to the Poisson or binomial laws in this case, or in cases where percentages are derived from the ratios of two *weights* or volumes, as in chemical analyses, or from an arbitrary observational scoring. With percentages, for example, it may be said, without describing the experimental variation in detail, that the variance must vanish at zero and 100 percent and is likely to be greatest in the middle. The formula $V = \lambda PQ$ is at least a first approximation to this situation. The angular transformation will approximately equalize a distribution of variances of this type, provided that λ is sufficiently small. We have, of course, returned to an "approximate" type of argument. It follows that the original data should be scrutinized carefully before deciding that a transformation is necessary and that any presumed opinions about the nature of the experimental variation should be verified as far as possible.

7. Summary. This paper discusses the theoretical basis for the use of the square root and inverse sine transformations in analyzing data whose experimental errors follow the Poisson and binomial frequency laws respectively.

The maximum likelihood equations of estimation are developed for each case, but are in general too complicated for frequent use. If, however, the expected yield of any plot is assumed to be an additive function of the treatment and soil effects in the transformed scale, a transformation can be found so that the equations of estimation assume the simple "normal theory" form. The transforms are closely related to the square roots and inverse sines respectively.

The nature of the assumed formula for the expected values is briefly discussed, and a χ^2 test is developed for the combined hypotheses that the prediction formula is satisfactory and that the experimental errors follow the assumed law.

Numerical examples are worked for both types of transformation. These indicate that even for data derived from small numbers, the square roots or inverse sines are good estimates of the correct transforms on almost all plots, except those which give zero yields in the Poisson case, or percentages near zero or 100 in the binomial case.

In practice, these new methods are not recommended to supplant the simple transformations for general use, because it can seldom be assumed that the whole of the experimental error variation follows the Poisson or binomial laws. The more exact analysis may, however, be useful (i) for cases in which the plot yields are very small integers or the ratios of very small integers (ii) in showing how to give proper weight to an occasional zero plot yield.

REFERENCES

- [1] W. G. COCHRAN, "Some difficulties in the statistical analysis of replicated experiments," *Empire J. Expt. Agric.*, Vol. 6 (1938), pp. 157-75.
- [2] M. S. BARTLETT, "The square root transformation in the analysis of variance," *J. Roy. Stat. Soc. Suppl.*, Vol. 3 (1936), pp. 68-78.
- [3] C. I. BLISS, "The transformation of percentages for use in the analysis of variance," *Ohio J. Sci.*, Vol. 38 (1938), pp. 9-12.
- [4] A. CLARK AND W. H. LEONARD, "The analysis of variance with special reference to data expressed as percentages," *J. Amer. Soc. Agron.*, Vol. 31 (1939), pp. 55-56.
- [5] W. G. COCHRAN, "The χ^2 distribution for the Binomial and Poisson series, with small expectations," *Ann. Eugen.*, Vol. 7 (1936), pp. 207-17.
- [6] P. V. SUKHATME, "On the distribution of χ^2 in samples of the Poisson series," *J. Roy. Stat. Soc. Suppl.*, Vol. 5 (1938), pp. 75-9.
- [7] C. I. BLISS, "The determination of the dosage-mortality curve from small numbers," *Quart. J. Pharmacy and Pharmacology*, Vol. 11 (1938), pp. 192-216.
- [8] A. W. JONES, "Practical field methods of sampling soil for wireworms," *J. Agric. Res.*, Vol. 54 (1937), pp. 123-34.
- [9] W. G. COCHRAN, "The information supplied by the sampling results," *Ann. App. Biol.*, Vol. 25 (1938), pp. 383-9.
- [10] R. A. FISHER AND F. YATES, *Statistical tables for agricultural, biological and medical research*, Edinburgh, Oliver and Boyd, 1938.
- [11] C. I. BLISS, "The analysis of field experimental data expressed in percentages," *Plant Protection* (Leningrad), 1937, pp. 67-77.
- [12] L. A. CARRUTH, "Experiments for the control of larvae of *Heliothis Obsoleta* Fabr.," *J. Econ. Ent.*, Vol. 29 (1936), pp. 205-9.
- [13] M. S. BARTLETT, "Some examples of statistical methods of research in agriculture and applied biology," *J. Roy. Stat. Soc. Suppl.*, Vol. 4 (1937), p. 168, footnote.

IOWA STATE COLLEGE,
AMES, IOWA

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

ORTHOGONAL POLYNOMIALS APPLIED TO LEAST SQUARE FITTING OF WEIGHTED OBSERVATIONS

BY BRADFORD F. KIMBALL

1. Introduction. Let the independent variable be denoted by x , and let it range over n consecutive integral values x_1 to x_n . Thus x represents the index-number of the ordered intervals at which observations are taken, where the intervals are all of equal length, and an index-number is assigned in consecutive order to every interval within the range of investigation, *whether observations occur in that interval or not*. Let y_x denote the observation measure (usually referred to as observed value), if such observation exists. Let w_x denote the weight of that observation, with weight zero assigned where observations are lacking.

To shorten the notation, summation over all values of x from x_1 to x_n will be denoted by the sign Σ . If a subscript and superscript is used, the context will indicate the variable to which the summation refers. The r th binomial coefficient will be denoted by $\binom{x}{r}$.

A system of polynomials $\phi_r(x)$, $r = 0, 1, 2, 3, \dots$ of degree r in x is said to be an orthogonal system, for the purposes of this paper, if they satisfy the relations

$$(1) \quad \sum W_x \phi_r(x) \phi_s(x) \quad \begin{cases} = 0, & r \neq s \\ \neq 0, & r = s. \end{cases}$$

To construct the polynomials, one may write them in the form

$$(2) \quad \begin{aligned} \phi_0(x) &= f_0(x) = \text{constant} \\ \phi_r(x) &= f_r(x) - \sum_{i=0}^{r-1} h_i \phi_i(x) \quad r = 1, 2, 3, \dots \end{aligned}$$

where the h_i are constants and the $f_r(x)$ are arbitrary polynomials of degree r . It then follows from the conditions of orthogonality that

$$(3) \quad h_i = \frac{\sum w_x f_r(x) \phi_i(x)}{\sum w_x [\phi_i(x)]^2}.$$

Thus when the polynomials $f_r(x)$ have been chosen for all r , the system of orthogonal polynomials for a given set of weights can be constructed and is uniquely determined except for a constant factor [1].

By virtue of the relation (2) and the conditions of orthogonality (1), it follows that

$$(4) \quad \Sigma w_x [\phi_r(x)]^2 = \Sigma w_x f_r(x) \phi_r(x).$$

Define the function $\Phi(r, k)$ by

$$(5) \quad \Phi(r, k) = \Sigma w_x f_r(x) \phi_k(x), \quad r = 0, 1, 2, 3, \dots$$

It follows from the relations (2) and (3) that

$$(6) \quad \phi_r(x) = f_r(x) - \sum_{i=0}^{r-1} \frac{\Phi(r, i)}{\Phi(i, i)} \phi_i(x)$$

where it is to be noted that this summation is independent of x .

Define q_r and Y_r by

$$(7) \quad q_r = \Sigma w_x [\phi_r(x)]^2 = \Sigma w_x f_r(x) \phi_r(x) = \Phi(r, r),$$

$$(8) \quad Y_r = \Sigma w_x y_x \phi_r(x).$$

Then if $u_r(x)$ represents the polynomial solution of degree r of the normal equations set up for observed values y_x and weights w_x ,

$$(9) \quad u_r(x) = \frac{Y_0}{q_0} + \frac{Y_1}{q_1} \phi_1(x) + \frac{Y_2}{q_2} \phi_2(x) + \dots + \frac{Y_r}{q_r} \phi_r(x).$$

If E^2 denotes the weighted sum of the squares of the discrepancies between the ordinates $u_r(x)$ of the fitted curve and the observed values y_x , then [2],

$$(10) \quad E^2 = \Sigma w_x [u_r(x) - y_x]^2 = \Sigma w_x y_x^2 - \sum_{i=0}^r \frac{Y_i^2}{q_i}.$$

The practicability of the use of orthogonal polynomials is thus seen to depend upon whether the quantities $\Phi(r, k)$ and Y_r can be evaluated in a reasonably simple manner.

The thesis of this paper is that if $f_r(x)$ is taken as the binomial coefficient $\binom{x}{r}$, one can effectively apply the method of orthogonal polynomials. This is made possible by the use of factorial moments in conjunction with an adding machine that prints cumulative totals.

In treating the same problem Aitken sets up the normal equations in terms of factorials, but considers the explicit use of orthogonal polynomials impractical. He writes: "the arbitrary nature of the weights stands in the way of any analytical sophistication; orthogonal polynomials emerge, but are not of great use; and the necessity of solving the moment equations cannot be circumvented" [3]. He prefers a determinantal method of solution of the normal

equations which the writer has found to be more involved from a practical point of view, than the present method, although it is elegant from a theoretical standpoint.

Thus although the present method is not new from the point of view of theory, the writer has found that forms made up by the use of the technique suggested below, offer an effective method for fitting polynomial curves to weighted observations.

2. Simplification of the problem when $f_r(x) = \binom{x}{r}$. Factorial moments S_r and M_r are defined by

$$(11) \quad S_r = \sum \binom{x}{r} w_x, \quad M_r = \sum \binom{x}{r} w_x y_x \quad r = 0, 1, 2, \dots$$

These moments are not difficult to compute and are readily checked as computed. Formula for $\Phi(r, k)$ then becomes

$$(12) \quad \Phi(r, k) = \sum \binom{x}{r} w_x \phi_k(x).$$

Thus since $\phi_0(x) = 1$, $\Phi(r, 0) = \sum \binom{x}{r} w_x = S_r$ and hence

$$\phi_1(x) = \binom{x}{1} - \frac{\Phi(1, 0)}{\Phi(0, 0)} = x - \frac{S_1}{S_0}.$$

Again

$$\begin{aligned} \Phi(r, 1) &= \sum \binom{x}{r} w_x \left(x - \frac{S_1}{S_0} \right) = \sum \binom{x}{r} \binom{x}{1} w_x - \frac{S_1}{S_0} \sum \binom{x}{r} w_x \\ &= (r+1)S_{r+1} + rS_r - \frac{S_1 S_r}{S_0}. \end{aligned}$$

Hence

$$q_1 = \Phi(1, 1) = 2S_2 + \left(1 - \frac{S_1}{S_0} \right) S_1.$$

A recursion formula for $\Phi(r, k)$ may be obtained by expanding $\phi_k(x)$ in formula (12) by means of (6). Thus

$$\begin{aligned} (13) \quad \Phi(r, k) &= \sum \binom{x}{r} \binom{x}{k} w_x - \sum_{i=0}^{k-1} \frac{\Phi(k, i)}{q_i} \left[\sum \binom{x}{r} w_x \phi_i(x) \right] \\ &= \sum \binom{x}{r} \binom{x}{k} w_x - \sum_{i=0}^{k-1} \frac{\Phi(r, i) \Phi(k, i)}{q_i}. \end{aligned}$$

The first term can be easily expressed as a linear combination of binomial coefficients, and thus as a linear combination of moments S_i .

The formula for Y_r can be broken down as follows:

$$\begin{aligned} Y_0 &= \sum w_x y_x = M_0, \\ (14) \quad Y_r &= \sum w_x y_x \phi_r(x) = \sum w_x y_x \binom{x}{r} - \sum_{i=0}^{r-1} \frac{\Phi(r, i)}{q_i} [\sum w_x y_x \phi_i(x)] \\ &= M_r - \sum_{i=0}^{r-1} \frac{\Phi(r, i)}{q_i} Y_i. \end{aligned}$$

Thus

$$\begin{aligned} Y_1 &= M_1 - \frac{S_1}{S_0} Y_0, \\ Y_2 &= M_2 - \frac{\Phi(2, 1)}{q_1} Y_1 - \frac{\Phi(2, 0)}{q_0} Y_0, \text{ etc.} \end{aligned}$$

3. General technique of computation. In determining the best fitting polynomial of degree r , the ratios $\Phi(r, i)/q_i$ are seen to play an important part. In a form for calculation, these quantities should receive simple designations such as b_i for a second degree curve, c_i for a third degree curve, etc. Suppose they are designated by R_i for a curve of degree r ; then

$$(15) \quad \phi_r(x) = \binom{x}{r} - \sum_{i=0}^{r-1} R_i \phi_i(x)$$

$$(16) \quad Y_r = M_r - \sum_{i=0}^{r-1} R_i Y_i$$

$$(17) \quad q_r = \sum \binom{x}{r}^2 w_x - \sum_{i=0}^{r-1} R_i \Phi(r, i)$$

and in determining $\Phi(r, k)$ for $k = 0, 1, 2, \dots, r-1$, formula (13) may be written:

$$(18) \quad \Phi(r, k) = \sum \binom{x}{r} \binom{x}{k} w_x - \sum_{i=0}^{k-1} R_i \Phi(r, i).$$

The fact that these quantities R_i appear as multipliers in so many of the fundamental formulas greatly simplifies the mechanics of the calculation, especially when a calculating machine is used.

In final determination of polynomial curve the differences of the polynomial at $x = 0$ are readily determined since the leading term of each orthogonal polynomial is a binomial coefficient and thus

$$\begin{aligned} (19) \quad \Delta^k \phi_r(0) &= - \sum_{i=0}^{r-1} R_i \Delta^k \phi_i(0), & k &= 1, 2, 3, \dots, r-1 \\ \Delta^r \phi_r(0) &= 1. \end{aligned}$$

Since the effectiveness of the method depends upon the availability of an adding machine which records a cumulative subtotal, the determination of the curve from the differences at the point $x = 0$ is not a hardship and indeed affords a quick and accurate means of setting up the curve for purposes of plotting and checking.

$$\begin{aligned}
 u_r(0) &= \frac{Y_0}{q_0} + \frac{Y_1}{q_1} \phi_1(0) + \frac{Y_2}{q_2} \phi_2(0) + \dots + \frac{Y_r}{q_r} \phi_r(0), \\
 (20) \quad \Delta^k u_r(0) &= \frac{Y_k}{q_k} + \frac{Y_{k+1}}{q_{k+1}} \Delta^k \phi_{k+1} + \dots + \frac{Y_r}{q_r} \Delta^k \phi_r(0), \\
 \Delta^r u_r(0) &= \frac{Y_r}{q_r}.
 \end{aligned}$$

The advantage of the use of orthogonal polynomials becomes particularly apparent when error formulae are to be used. The formula for the sum of the squares of the discrepancies, denoted by E^2 , is given above (formula (10)). The estimated variance V of the weighted observations about the fitted curve is thus $E^2/(n - r - 1)$ where n is the number of values of x used in fitting and r is the degree of the curve fitted. Recalling that the matrix of the normal equations is of the diagonal form with diagonal elements q_0, q_1, \dots, q_r it follows that the coefficient Y_k/q_k of $\phi_k(x)$ in the expansion of $u_r(x)$ has the variance V/q_k .

Furthermore the variance of the ordinate of the fitted curve $u_r(x)$ at a point x due to sampling variations in the determination of the coefficients of the curve, under the assumption that the weights and values of the independent variable x do not involve errors, has the simple form

$$(21) \quad \begin{array}{l} \text{Variance of } u_r(x) \\ \text{at point } x = V \left[\frac{\phi_0^2(x)}{q_0} + \frac{\phi_1^2(x)}{q_1} + \dots + \frac{\phi_r^2(x)}{q_r} \right] \end{array}$$

since the covariances of the orthogonal polynomials are zero [4].

REFERENCES

- [1] G. SZEGO, *Orthogonal Polynomials*, Colloquium Publications, Amer. Math. Soc., New York, 1939, Vol. 23, Chapter II.
- [2] MAX SASULY, *Trend Analysis of Statistics*, The Brookings Institution, Washington, 1934, pp. 296-297.
- [3] A. C. AITKEN, "On Fitting Polynomials to Weighted Data by Least Squares," *Proc. Roy. Soc. of Edinburgh*, Session 1933-34, Vol. 54, Part I, No. 1, p. 2.
- [4] HENRY SCHULTZ, "The Standard Error of a Forecast from a Curve," *Jour. Amer. Stat. Assn.*, June 1930; or Whittaker & Robinson, *The Calculus of Observations*, Blackie & Son Ltd., London, 1937, 2d ed., p. 242.

PORT WASHINGTON, N. Y.

**COMBINATORIAL FORMULAS FOR THE r th STANDARD MOMENT
OF THE SAMPLE SUM, OF THE SAMPLE MEAN,
AND OF THE NORMAL CURVE**

By P. S. DWYER

The standard moments of the normal curve are usually expressed by the two statements [1, p. 97]

$$(1) \quad \begin{cases} \alpha_{2s} &= \frac{(2s)!}{2^s s!} \\ \alpha_{2s+1} &= 0 \end{cases}$$

It is of some interest to note that these two statements may be generalized into a single statement by observing that $\frac{(2s)!}{2^s s!}$ is the number of ways in which $2s$ things can be grouped in pairs and that 0 is the number of ways in which $2s + 1$ things can be grouped in pairs. It is obvious that an odd number of things can not be grouped in pairs since there must be at least one unpaired unit. It is clear, too, that the number of orders in which $2s$ things can be grouped in pairs is $\binom{2s}{2} \binom{2s-2}{2} \binom{2s-4}{2} \cdots \binom{4}{2} \binom{2}{2}$ and this is $\frac{(2s)!}{2^s}$. However if the resulting paired groups (rather than the orders of grouping) are counted it is seen that each paired grouping is repeated $s!$ times so that $\frac{(2s)!}{2^s s!}$ represents the number of ways $2s$ things can be grouped in pairs. If we arbitrarily define the number of ways 0 things can be grouped in pairs to be 1 (or if we limit our theorem to values of $r > 0$) we may say "The r th standard moment of the normal curve is equal to the number of ways in which r things can be grouped in pairs."

As presented above the combination representation is used primarily as a means of unification of results. However, it is possible to derive the standard moments of the normal curve in such a way as to indicate the term $\frac{(2s)!}{2^s s!}$ early in the proof and to trace it throughout the proof. I follow the method outlined by H. C. Carver [2] in obtaining the normal distribution as the limit of the distribution of sample sums (or of sample means) though I use a somewhat different notation [3, p. 5]. If we let $\binom{1^r}{p_1^{r_1} \cdots p_s^{r_s}}$ represent the number of ways in which r units can be collected with π_1 groups containing p_1 units, π_2 groups containing p_2 units, etc., then the multinomial theorem can be expressed as [3, p. 17]

$$(2) \quad (1)^r = \sum \binom{1^r}{p_1^{r_1} \cdots p_s^{r_s}} (p_1^{r_1} \cdots p_s^{r_s})$$

where the summation is taken over all possible partitions $p_1^{r_1} \dots p_s^{r_s}$ of r and the expression $(p_1^{r_1} \dots p_s^{r_s})$ represents the power product form [3, p. 14] which is $\pi_1! \pi_2! \dots \pi_s!$ times the monomial symmetric function. If ρ represents the number of parts of the partition then

$$\rho = \pi_1 + \pi_2 + \dots + \pi_s$$

while

$$r = p_1 \pi_1 + p_2 \pi_2 + \dots + p_s \pi_s.$$

Now it can be shown from (2) in the case of infinite sampling that

$$(3) \quad \bar{\mu}_{r:(1)} = \sum \left(p_1^{r_1} \dots p_s^{r_s} \right) n^{(\rho)} (\bar{\mu}_{p_1})^{r_1} \dots (\bar{\mu}_{p_s})^{r_s}$$

and since $\bar{\mu}_1 = 0$, it is only necessary to sum over all partitions which have no unit part. We have then, dividing by $[\bar{\mu}_{2:(1)}]^{1r} = [n\bar{\mu}_2]^{1r}$

$$(4) \quad \alpha_{r:(1)} = \sum \left(p_1^{r_1} \dots p_s^{r_s} \right) \frac{n^{(\rho)}}{n^{1r}} (\alpha_{p_1})^{r_1} \dots (\alpha_{p_s})^{r_s}.$$

We have now a formula for the r th standard moment of the sample sum which is expressed essentially in combination notation since the quantity $\left(p_1^{r_1} \dots p_s^{r_s} \right)$ represents the number of ways in which r units can be grouped to form π_1 groups containing p_1 units, π_2 groups containing p_2 units, etc. All non-unitary groupings of r are formed, each combinatorial coefficient is computed and multiplied by $n^{(\rho)}/n^{1r}$ times the product of the corresponding α 's, and the sums are formed. It might be noted that the formula for the r th standard moment of the sample mean is identical with (4) while the corresponding finite sampling (without replacements) formula is

$$(5) \quad \alpha_{r:(1)} = \sum \left(p_1^{r_1} \dots p_s^{r_s} \right) \frac{N^{\rho} P_{p_1^{r_1} \dots p_s^{r_s}}}{N^{1r} P_2^{1r}} (\alpha_{p_1})^{r_1} \dots (\alpha_{p_s})^{r_s}.$$

The P 's are defined in previous papers [2, p. 105-6][3, p. 113].

We obtain the formula for the r th standard moment of the normal curve by taking the limit of (4) as $n \rightarrow \infty$. (H. C. Carver has pointed out [2, p. 121] that this method of derivation imposes fewer restrictions than does the derivation from Hagen's hypothesis.) Each partition term will approach zero as n approaches infinity if $\rho < \frac{1}{2}r$. Now the only non-unitary partition in which ρ is not less than $\frac{1}{2}r$ is the partition 2^{1r} and we can have this partition only when r is even. Now the limit as n approaches infinity of $n^{(\rho)}/n^{1r}$ is unity and we have, in the limiting case

$$(6) \quad \alpha_r = \begin{cases} \left(\frac{1^r}{2^{1r}} \right) & \text{if } r \text{ is even.} \\ 0 & \text{if } r \text{ is odd.} \end{cases}$$

Since $\binom{1^r}{2^r}$ is the number of ways r units can be grouped in pairs when r is even and since 0 is the number of ways r units can be grouped in pairs where r is odd, it follows that the r th standard moment of the normal curve is the number of ways in which r units can be grouped in pairs.

This development is of interest in that it makes possible the tracing of the value $\binom{1^r}{2^r}$ back through the various stages of the development to the coefficient of (2^r) in the power product expansion of the multinomial theorem.

REFERENCES

- [1] H. C. CARVER, "Frequency Curves," Chapter VII of *Handbook of Mathematical Statistics*.
- [2] H. C. CARVER, "Fundamentals of the theory of sampling," *Annals of Math. Stat.*, Vol. 1 (1930), pp. 101-121.
- [3] P. S. DWYER, "Combined Expansions of Products of Symmetric Power Sums and of Sums of Symmetric Power Products with Application to Sampling," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 1-47, 97-132.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICHIGAN

ON A METHOD OF SAMPLING¹

BY E. G. OLDS

It is recorded that Diogenes fared forth with a lantern in his search for an honest man. History does not tell us how many dishonest men he encountered before he found the first honest one but, judging from the fact that he took his lantern, apparently he expected to have a long search. The general problem of sampling inspection, of which the above is a special case, can be stated as follows:

Given a lot, of size m , containing s items of a specified kind. If items are to be drawn without replacement until i of the s items have been drawn, how many drawings, on the average, will be necessary?

Uspensky² has solved a problem concerning balls in an urn, from which the answer to the above question can be obtained for the special case $i = 1$. For the general case, the distribution for the number n of the drawing in which the i th specified item appears, is given by terms of the series:

$$(1) \quad \nu'_0 = \sum_{n=1}^{m-s+i} \frac{C_{n-1,i-1} C_{m-n,s-i}}{C_{m,s}} = \sum_{n=0}^{\infty} \frac{C_{n-1,i-1} C_{m-n,s-i}}{C_{m,s}},$$

¹ Presented to The Institute of Mathematical Statistics, Dec. 27, 1938, at Detroit, Mich., as part of a paper, entitled "Remarks on two methods of sampling inspection."

² J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937, p. 178.

where the first symbol indicates the number of ways of choosing $i - 1$ of the specified items to fill the first $n - 1$ places, the second symbol indicates the number of ways of disposing of $s - 1$ specified items in the last $m - n$ places, and the denominator gives the number of ways that the s items can be scattered through the lot. In order to get the average number of draws we multiply ν'_0 by n and sum. Then we have

$$(2) \quad \nu'_1 = \sum_{n=0}^{\infty} \frac{n C_{n-1, i-1} C_{m-n, s-1}}{C_{m, s}} = \frac{i(m+1)}{s+1} \sum_{n=0}^{\infty} \frac{C_{n, i} C_{m-n, s-1}}{C_{m+1, s+1}} = \frac{i(m+1)}{s+1}.$$

Example 1. On a table of 200 bargain shirts there are 5 which have a 15 in. neckband and 35 in. sleeves. How many shirts must be examined, on the average, to find two of the desired kind?

Solution. For this case, $m = 100$, $s = 5$, $i = 2$. Therefore $\bar{n} = [2(201)] \div 6 = 67$. Thus, an average of 67 shirts must be examined.

Suppose μ_K represents the K th moment about the mean, ν_K the K th moment about the origin, and ν'_K the moment relation given by

$$(3) \quad \nu'_K = (\nu_1 + K - 1)^{(K)},$$

where $(\nu_1 + K - 1)^{(K)}$ represents the result of expanding $(\nu + K - 1)^{(K)}$ and changing the exponent of ν to the corresponding subscript. (For example, $\nu'_3 = (\nu_1 + 2)^{(3)} = \nu_3 + 3\nu_2 + 2\nu_1$.) It is easy to derive the recurrence relation

$$(4) \quad \nu'_K = \frac{(i + K - 1)(m + K)}{s + K} \nu'_{K-1}.$$

From this result the computation of the moments about the mean is theoretically direct. Actually the results do not seem to be very compact. The variance is given by

$$(5) \quad \mu_2 = \frac{(m+1)(m-s)}{(s+1)^2(s+2)} [i(s+1) - i^2].$$

In case s is unknown and n is known for a particular value of i , we may estimate s , (or rather $\frac{1}{s+1}$), by using the relation, $n = \frac{i(m+1)}{s+1}$. Then

$$(6) \quad \frac{1}{s+1} \text{ est.} = \frac{n}{i(m+1)},$$

and the variance, using this estimate, is given by

$$(7) \quad \text{Variance of } \left(\frac{1}{s+1} \right) \text{ est.} = \frac{n}{n + i(m+1)} \cdot \frac{1}{i(m+1)} \left[\frac{n}{i} - 1 \right] \left[1 - \frac{n}{m+1} \right].$$

Example 2. In order to check a box of 144 screws, screws are drawn until 10 good screws are obtained. In a particular case only 10 drawings were necessary. Estimate the number of good screws in the lot.

Solution. Here $m = 144$, $i = 10$, $n = 10$. The estimate for s is obtained

from $\left(\frac{1}{s+1}\right)$ est. = $\frac{10}{10(145)} = \frac{1}{145}$ and, as might be expected, the conclusion is that all the screws are good. Furthermore the variance of the estimated quantity is zero.

It is obvious that the number of draws necessary to obtain any particular number of specified items is correlated with the numbers of draws for lesser numbers of items. To investigate this, let us suppose that n_j represents the number of draws to obtain exactly j specified items and that $x_j = n_j - n_{j-1}$. It follows immediately from our previous results, that

$$(8) \quad E(x_1) = E(x_2) = E(x_3) = \dots = \frac{m+1}{s+1}.$$

This result could be obtained from the fact that, corresponding to any arrangement of the lot for which $x_a = a$ and $x_b = b$, there is another arrangement where $x_a = b$ and $x_b = a$, formed by moving $a - b$ of the non-specified items from the first group to the second. From this fact we see, also, that

$$(9) \quad E(x_1^2) = E(x_2^2) = E(x_3^2) = \dots$$

But $x_1 = n_1$ and $\sigma_{n_1}^2 = \frac{(m+1)(m-s)}{(s+1)^2(s+2)} [s+1-1] = ds$.

Therefore,

$$(10) \quad \sigma_{x_1}^2 = \sigma_{x_2}^2 = \dots = ds.$$

But, from our previous formula we have

$$\sigma_{n_2}^2 = d(2s-2), \quad \sigma_{n_3}^2 = d(3s-6), \text{ etc.}$$

Since $n_2 = x_1 + x_2$, it follows that

$$\sigma_{n_2}^2 = \sigma_{x_1}^2 + 2r_{x_1, x_2} \sigma_{x_1} \sigma_{x_2} + \sigma_{x_2}^2$$

where r_{x_1, x_2} is the correlation between x_1 and x_2 . Therefore,

$$(11) \quad r_{x_1, x_2} = -1/s.$$

Also, since $x_1 = n_2 - x_2$, it follows that

$$(12) \quad r_{n_2, x_2} = \sqrt{\frac{s-1}{2s}}.$$

Likewise, from $x_2 = n_2 - x_1$, we get

$$(13) \quad r_{n_2, x_1} = \sqrt{\frac{s-1}{2s}}.$$

Finally, we obtain the three general results

$$(14) \quad r_{n_i, x_{i+1}} = -\sqrt{\frac{i}{s(s-i+1)}},$$

$$(15) \quad r_{n_i, x_i} = \sqrt{\frac{s-i+1}{si}},$$

$$(16) \quad r_{n_{i+1}, n_i} = \sqrt{\frac{i(s-i)}{(i+1)(s-i+1)}}.$$

Example 3. The cards of a deck are turned one by one until two aces have appeared. The second ace appears when the 36th card is turned. How many more cards should one expect to have to turn to find a third ace?

Solution. Here $m = 52$, $s = 4$, $i = 2$, $n_2 = 36$.

Then $\bar{n}_2 = 2 \cdot \frac{53}{5}$, $\bar{x}_3 = \frac{53}{5}$, and $r_{n_2, x_3} = -\sqrt{\frac{2}{4(4-2+1)}} = -\frac{\sqrt{6}}{6}$. Also

$\sigma_{x_3} = \sqrt{4d}$ and $\sigma_{n_2} = \sqrt{6d}$. Since $\frac{x_3 - \bar{x}_3}{\sigma_{x_3}} = r_{n_2, x_3} \frac{(n_2 - \bar{n}_2)}{\sigma_{n_2}}$, we have

$$x_3 = \frac{53}{5} - \frac{2}{\sqrt{6}} \cdot \frac{\sqrt{6}}{6} \left(36 - \frac{106}{5} \right) = \frac{17}{3}.$$

Of course this result could have been obtained more directly by noting that there were two aces left among the 16 remaining cards.

Conclusion. The results given in this note might be useful when it is necessary to estimate the number of items to be drawn in order to secure a desired number of a particular type, such as may be the case in obtaining a sample with previously defined characteristics. Also the note disproves such intuitive notions as the one that when looking for a desired record, one is most likely to have to search the whole pile to find it. As far as methods of sampling inspection are concerned, the one implied in this note has little to recommend it.

CARNEGIE INSTITUTE OF TECHNOLOGY,
PITTSBURGH, PA.

RANK CORRELATION WHEN THERE ARE EQUAL VARIATES¹

BY MAX A. WOODBURY

If there is given a set of number pairs

$$(1) \quad (X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N),$$

we may assign to each variate its "rank" (i.e. one more than the number of corresponding variates in the set greater than the given variate). In this way there is obtained a set of pairs of ranks

$$(2) \quad (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N).$$

¹ Presented at the fall meeting, Mich. section of the Math. Assn. of America, Nov. 18, 1939, Kalamazoo College.

If we assume that $X_i \neq X_j$ and $Y_i \neq Y_j$ when $i \neq j$ then it follows that each integer from 1 to N appears once and only once in the x 's and the same holds for the y 's. This leads at once to the formulas:

$$(3a) \quad \sum_{i=1}^N x_i = \sum_{i=1}^N y_i = \sum_{i=1}^N i = N(N+1)/2,$$

$$(3b) \quad \sum_{i=1}^N x_i^2 = \sum_{i=1}^N y_i^2 = \sum_{i=1}^N i^2 = N(N+1)(2N+1)/6.$$

When these results are substituted in the expression for the product moment correlation coefficient we have after simplifying [1],

$$(4) \quad \rho = 1 - 6 \sum_{i=1}^N D_i^2 / N(N^2 - 1) \quad \text{where } D_i = x_i - y_i.$$

If we consider the case of equal variates and follow the rule for assigning ranks given in the first paragraph, the resulting method is known as the bracket-rank method. The use of (4) in the calculation of ρ by this method is not strictly valid, because not every integer appears in the summations and so neither (3a) nor (3b) is true.

The more accurate mid-rank method assigns to each of the equal variates the average of the ranks that would be assigned if we were to give them an arbitrary order. This method preserves (3a) but not (3b). In this paper ρ_M indicates the value of ρ as calculated by (4) when the mid-rank method is used.

In a method due to DuBois [2], the equal variates are assigned the same rank so as to satisfy (3b). In this case (3a) is not satisfied.

If we assign the ranks to the equal variates in an arbitrary way, then (3a) and (3b) are of course satisfied and the use of (4) is valid. There are two disadvantages to such a method; first, the equal variates are treated differently, and second, the assignment of ranks is arbitrary. These difficulties are removed if one uses the average of the values of ρ corresponding to all possible ways of arbitrarily assigning ranks to the equal variates. Since ρ is linear in $\sum_i D_i^2$ the average value of ρ may be obtained from the average value of $\sum_i D_i^2$ and the use of (4).

Let us first consider the simple case of two equal variates in one of the variables, say X . It is clear that there are only two possible ways of assigning ranks, and that if we arrange the series by the assigned x ranks, the resulting series differ only in the y ranks corresponding to the equal X variates. If we denote the two x ranks to be assigned by m and $m+1$ and the y 's corresponding for a particular arrangement by y_m and y_{m+1} we have for the average $\sum_i D_i^2$ the expression

$$(5a) \quad \sum_{x=1}^{m-1} (x - y_x)^2 + \sum_{x=m+2}^N (x - y_x)^2 \\ + \frac{1}{2}[(m - y_m)^2 + (m + 1 - y_{m+1})^2 + (m - y_{m+1})^2 + (m + 1 - y_m)^2].$$

By the mid-rank method the corresponding expression is

$$(5b) \quad \sum_{z=1}^{m-1} (x - y_z)^2 + \sum_{z=m+2}^N (x - y_z)^2 + (m + \frac{1}{2} - y_m)^2 + (m + \frac{1}{2} - y_{m+1})^2.$$

The correction Δ_2 to be added to the mid-rank $\sum_i D_i^2$ to get the average $\sum_i D_i^2$ is, by subtracting (5b) from (5a) and simplifying,

$$(6) \quad \Delta_2 = \frac{1}{2}.$$

To get Δ_K in the more general case of several equal variates, we need only consider the difference between the average value of $\sum_i D_i^2$ and that obtained by the mid-rank method. If there are K equal X variates we may assign the ranks in $K!$ ways, this results in $K!$ permutations of the y ranks for the sets arranged in order of their assigned x ranks. In $(K-1)!$ permutations y_{m+i} corresponds to the x rank of $m+i$ so that the correction to the mid-rank $\sum_{i=1}^N D_i^2$ is

$$(7) \quad \begin{aligned} \Delta_K &= \frac{(K-1)!}{K!} \left[\sum_{j=0}^{K-1} \sum_{i=0}^{K-1} (m+i-y_{m+i})^2 \right] - \sum_{j=0}^{K-1} \left(m + \frac{K-1}{2} - y_{m+j} \right)^2 \\ &= \frac{1}{K} \sum_{j=0}^{K-1} \sum_{i=0}^{K-1} \left[(m+i-y_{m+i})^2 - \left(m + \frac{K-1}{2} - y_{m+j} \right)^2 \right] = \frac{K(K^2-1)}{12}. \end{aligned}$$

It is to be noticed that the correction is positive and depends *only* on the number of equal X variates. From this it can be concluded that for more than one group of equal variates no matter whether X 's or Y 's we can obtain the average $\sum_i D_i^2$ by computing a correction for each group and then adding these corrections to get the total correction to the mid-rank $\sum_i D_i^2$. Then as before noted we can by (4) calculate the average ρ (denoted as $\bar{\rho}$).

This correction to $\sum_i D_i^2$ may be converted into a correction to ρ_M . That is

$$\text{if } \delta_{N,K_i} = \frac{6\Delta_{K_i}}{N(N^2-1)} = \frac{K_i(K_i^2-1)}{2N(N^2-1)}, \text{ then}$$

$$(8) \quad \bar{\rho} = \rho_M - \sum_i \delta_{N,K_i},$$

where the summation extends over all groups of equal variates, and K_i is the number of equal variates in the i th group.

A table of $\delta_{N,K}$ for different values of N and K is given, and also a table of Δ_K . The values Δ_K are given in the top row of the table, while the $\delta_{N,K}$ are given in the rows below.

Table of Δ_K and δ_{NK}

$N \backslash K$	2	3	4	5	6	7	8	9	10	11	12	13
Δ_K	0.5000	2.000	5	10	17.5	28	42	60	82.5	110	143	182
δ_{NK}												
3	1250	—	—	—	—	—	—	—	—	—	—	—
4	0500	2000	—	—	—	—	—	—	—	—	—	—
5	0250	1000	2500	—	—	—	—	—	—	—	—	—
6	0143	0571	1429	2857	—	—	—	—	—	—	—	—
7	0089	0357	0893	1786	3125	—	—	—	—	—	—	—
8	0060	0238	0595	1190	2083	3333	—	—	—	—	—	—
9	0042	0166	0417	0833	1458	2333	3500	—	—	—	—	—
10	0030	0121	0303	0606	1061	1697	2546	3636	—	—	—	—
11	0023	0091	0227	0455	0795	1273	1909	2727	3750	—	—	—
12	0017	0070	0175	0350	0612	0979	1469	2098	2885	3846	—	—
13	0014	0055	0137	0275	0480	0769	1154	1648	2266	3022	3929	—
14	0011	0044	0110	0220	0385	0615	0923	1319	1813	2418	3143	4000
15	0009	0036	0089	0179	0313	0500	0750	1071	1473	1964	2554	3250
16	0007	0029	0074	0147	0257	0412	0618	0882	1213	1618	2103	2676
17	0006	0025	0061	0123	0214	0343	0515	0735	1011	1348	1752	2230
18	0005	0021	0052	0103	0181	0289	0433	0619	0851	1135	1476	1878
19	0004	0018	0044	0088	0154	0246	0368	0526	0724	0965	1254	1596
20	0004	0015	0038	0075	0132	0211	0316	0451	0620	0827	1075	1368
21	0003	0013	0032	0065	0114	0182	0273	0390	0536	0714	0929	1182
22	0003	0011	0028	0056	0099	0158	0237	0339	0466	0621	0807	1028
23	0002	0010	0025	0049	0086	0138	0208	0296	0408	0543	0708	0899
24	0002	0009	0022	0043	0076	0122	0183	0261	0359	0478	0622	0791
25	0002	0008	0019	0038	0067	0108	0162	0231	0317	0423	0550	0700
26	0002	0007	0017	0034	0060	0096	0144	0205	0282	0376	0489	0622
27	0002	0006	0015	0031	0053	0085	0128	0183	0252	0336	0437	0556
28	0001	0005	0014	0027	0048	0077	0115	0164	0226	0301	0391	0498
29	0001	0005	0012	0025	0043	0069	0103	0148	0203	0271	0352	0448
30	0001	0004	0011	0022	0039	0062	0093	0133	0184	0245	0318	0405
35	0001	0003	0007	0014	0025	0039	0059	0084	0116	0154	0200	0255
40	0000	0002	0005	0009	0016	0026	0039	0056	0077	0103	0134	0171
45	0000	0001	0003	0007	0012	0018	0028	0040	0054	0072	0094	0120
50	0000	0001	0002	0004	0007	0011	0016	0023	0032	0043	0055	0070
60	0000	0001	0001	0003	0005	0008	0012	0017	0023	0031	0040	0051
70	0000	0000	0001	0002	0003	0005	0007	0010	0014	0019	0025	0032
80	0000	0000	0001	0001	0002	0003	0005	0007	0010	0013	0017	0021
90	0000	0000	0000	0001	0001	0002	0003	0005	0007	0009	0012	0015
100	0000	0000	0000	0000	0001	0002	0003	0004	0005	0007	0009	0011

As an example of the use of the table we will consider the following problem, [2, p. 56], with the ranks assigned as for the mid-rank method.

Subject	I	II
A	1	2.5
B	4	10
C	4	2.5
D	4	5
E	4	7
F	4	2.5
G	7	8
H	8	2.5
I	9.5	6
J	9.5	12
K	11	11
L	13	13
M	13	9
N	13	14

For the mid-rank method we have

$$\sum_{i=1}^{14} D_i^2 = 119.5, N = 14,$$

$$\rho_M = 1 - \frac{6(119.5)}{14(196 - 1)} = 0.7374.$$

Referring to the table we find that

K_i	ΔK_i	δ_{NK_i}
2	0.5	0.0011
3	2.0	0.0044
4	5.0	0.0110
5	10.0	0.0220
Total	17.5	0.0385

We know that $\bar{\rho} = 1 - \frac{6(119.5 + 17.5)}{14(196 - 1)} = 0.6989$ and in terms of δ_{NK_i}

$$\bar{\rho} = 0.7374 - 0.0385 = 0.6989$$

The value given by DuBois for his method is 0.7511.

Conclusion. A method has been developed for the treatment of rank correlation where there are groups of equal variates. The method consists of applying a generally small correction to the value as ordinarily calculated by the mid-rank method in order to find the value which would be obtained by averaging the values of the rank correlation coefficient for all possible ways of arbitrarily assigning ranks to the equal variates. Thanks are due Professor P. S. Dwyer, without whose aid and encouragement this paper would not have been written.

REFERENCES

- [1] YULE AND KENDALL, *Introduction to the Theory of Statistics*, p. 248. London, 1937.
- [2] P. DuBois, "Formulas and Tables for Rank Correlation," *Psychol. Rev.*, Vol. 3(1939) pp. 45-56.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICHIGAN

NOTE ON THEORETICAL AND OBSERVED DISTRIBUTIONS OF REPETITIVE OCCURRENCES

By P. S. OLMSTEAD

1. A simple problem of repetitive occurrences. Two questions which the engineer often desires to answer whenever he has a new type of apparatus or a new design of an old type of apparatus are: How many times will it perform its intended function without failure? and How many times will it fail to perform its intended function in a given length of time? To do this, he selects a number of what he believes to be identical units of the apparatus and gives each unit a performance test under a uniform test procedure. The number of satisfactory operations prior to the first observed failure to perform this operation is called a "run" and is a measure of the type desired for each unit.

If it is assumed that the probability of failure at any operation is a constant, q , and the probability of satisfactory operation is $1 - q$ or p , then the mathematical probability of runs of 0, 1, 2, 3 ... satisfactory operations for any unit are

$$(1) \quad q, pq, p^2q, p^3q, \dots$$

respectively.

Let x denote the number of satisfactory operations in any run. The mean value of x , say m_x , is given by

$$(2) \quad m_x = \frac{p}{q}.$$

The variance of x is

$$(3) \quad \sigma_x^2 = \frac{p}{q^2}.$$

The first step in practice is to determine whether there exists a constant probability, p , by means of the application of the operation of statistical control.¹ Expressions (1), (2), and (3) provide the necessary information for doing this. When a constant probability exists as evidenced by at least 25 consecutive samples of 4 units each the following practical procedure has been found to be satisfactory.

1. An estimate of p (or q), the sole parameter of the distribution, can be obtained from the average length of run in the sample. If p is less than 0.6 and if the sample size is large, a reasonably good estimate of p can be obtained from the proportion of the sample having runs of zero length.

2. The probability of getting runs of length x or more is p^x . Thus, if a minimum (or maximum) value of the probability, p^x , is chosen, a maximum

¹W. A. Shewhart, "Statistical Method from the Viewpoint of Quality Control," The Department of Agriculture Graduate School, Washington, 1939, Chapter I.

(or minimum) expected length of run can be computed for use as a criterion for looking for assignable causes of variation in the length of individual runs by using the estimated value of p .

3. The average and standard deviation to be used in calculating the limits to be applied to successive samples of rational sub-groups in accordance with the Shewhart² Criterion I are given by Equations (2) and (3) in which the estimates of p and q are substituted.

2. Application to a signal transmission problem. The theoretical solution given above is a direct answer to the first question at the head of this note.

TABLE I

Observed distributions of runs of x occurrences of event E for various test periods of apparatus life

No. of Occurrences per Period	Freq.	Test Period									
		1	2	3	4	5	6	7	8	11	15
x											
0	n_0	878	1519	961	723	541	407	343	266	160	77
1	n_1	77	226	207	206	171	148	129	97	70	35
2	n_2	2	31	44	55	68	46	52	39	37	27
3	n_3	1	3	8	18	15	19	13	22	19	10
4	n_4		2	1	2	—	6	5	5	7	3
5	n_5			—	1	1	3	1	1	5	2
6	n_6			1			1		—	1	2
7	n_7								1	—	—
8	n_8									2	1
Sample Size	n	958	1781	1222	1005	796	630	543	431	301	157

The second question is also of interest particularly when failure to perform an operation does not impair the apparatus unit for performance of additional operations. In cases of this type, the engineer often lets his test continue for test periods of particular lengths, measured in numbers of operations or sometimes in intervals of time (i.e., time intervals are often considered to be proportional to numbers of operations) and observes the number of failures during the test period for each unit. Thus, he may, after he has assured himself that control exists, arrange his data for each test period to show the frequency of occurrence of 0, 1, 2, 3, . . . failures per unit.

Data of this type which are typical of those found in other studies made

² Loc. cit.

during the past two years are presented in Table I. These were obtained in a signal transmission study in which the data for successive periods were obtained

TABLE II

Comparison of observed and theoretical values of averages and variances for distributions of Table I

Statistic or Parameter		Test Period									
		1	2	3	4	5	6	7	8	11	15
$\bar{q} = \frac{n_0}{n}$	observed	.916	.853	.786	.719	.679	.646	.632	.617	.532	.491
\bar{x}	observed	.098	.171	.269	.381	.448	.543	.537	.633	.917	1.026
$m_x = \frac{\bar{p}}{\bar{q}}$	theoretical*	.091	.172	.272	.390	.471	.548	.583	.620	.881	1.039
σ_x^2	observed	.091	.200	.343	.497	.556	.832	.760	1.075	1.783	1.921
$\sigma_x^2 = \frac{\bar{p}}{\bar{q}^2}$	theoretical*	.098	.202	.345	.542	.693	.848	.924	1.005	1.658	2.117

* Based on assumption that \bar{q} is the true value of q .

TABLE III

Theoretical distributions corresponding to distributions of Table I calculated by using $\bar{q} = \frac{n_0}{n}$ as the true value of q

No. of Occurrences per Period	Freq.	Test Period									
		1	2	3	4	5	6	7	8	11	15
x											
0	n_0^*	878.0	1519.0	961.0	723.0	541.0	407.0	343.0	266.0	160.0	77.0
1	n_1	73.3	233.5	205.3	202.8	173.3	144.1	126.4	101.9	74.9	39.2
2	n_2	6.1	32.9	43.8	56.9	55.5	51.0	46.6	39.0	35.1	20.0
3	n_3	.5	4.8	9.4	16.0	17.8	18.0	17.1	14.9	16.5	10.2
4	n_4	.1	.7	2.0	4.5	5.7	6.4	6.3	5.7	7.7	5.2
5	n_5		.1	.4	1.3	1.8	2.3	2.3	2.2	3.6	2.6
6	n_6			.1	.4	.6	.8	.9	.8	1.7	1.4
7	n_7				.1	.2	.3	.3	.3	.8	.7
8	n_8					.1	.1	.1	.1	.4	.3
9 or over	$n_{9-\infty}$.1	.3	.4
Sample Size	n^*	958	1781	1222	1005	796	630	543	431	301	157

* The observed values of n_0 and n form the basis for the calculated distributions.

for separate units. Since each set of these data passed the scrutiny for control, there is justification for assuming that a statistical universe exists and that its functional form may be derived from the observed distribution. It was found

that these data were consistent with the assumption that, where the probability of non-occurrence of a failure on a unit in the test period was q , the probability of exactly x failures on a unit was $p^x q$. This set of mathematical probabilities is shown in (1) with q redefined to apply in this case to non-occurrence of a failure.

Observed and "Theoretical" values of the averages and variances for the observed distributions are shown in Table II. The basis for calculating the theoretical values was to take the ratio (designated \bar{q}) of n_0 to n for each distribution as the estimate of the true value, q . Distributions as shown in Table III

TABLE IV

Test of fit of theoretical to observed distributions (Table III and Table I, respectively)

	Test Period									
	1	2	3	4	5	6	7	8	11	15
χ^2 *	2.24	0.20	0.32	2.09	9.79	0.65	3.20	6.27	1.07	3.98
Degrees of Freedom	1	2	2	3	3	3	3	3	4	4
P_{χ^2}	.13	.90	.87	.55	.02	.87	.36	.10	.90	.41

* Minimum number in cell for theoretical distribution taken as 5.

were calculated from each \bar{q} . These distributions were tested against the observed distributions by means of the χ^2 test with the results shown in Table IV, which are all within reasonable limits of what might be expected when a constant probability exists.

3. Conclusions. When a constant probability applies to each operation in a repetitive process this note shows how to establish criteria for identifying significantly long or short lengths for individual runs and significantly high or low average lengths for groups of several runs. A problem taken from the field of signal transmission gives assurance of the existence of this type of distribution in practice.

BELL TELEPHONE LABORATORIES,
NEW YORK, N. Y.

